Committee on Population

Workshop on Women’s Mental Health across the Life Course through a Sex-Gender Lens

March 7, 2018

The National Academies of

SCIENCE & ENGINEERING & MEDICINE

The District Architecture Center

421 7th Street, NW

Washington, DC 20004
# TABLE OF CONTENTS

## Tab A
1. Agenda
2. Guidance to Presenters
3. Steering Committee and Presenter Biosketches
4. Project Description
5. Committee on Population (CPOP) Roster

## Tab B
### The Big Picture: Life Course Framing, Population Patterns, Measurement, Methods (BACKGROUND READINGS)


## Tab C
### Psychological and Structural Factors in Women's Mental Health (BACKGROUND READINGS)

### TABLE OF CONTENTS

#### Tab D  Mental Health over the Life Course: Adolescence (BACKGROUND READINGS)


#### Tab E  Mental Health over the Life Course: Mid-Life (BACKGROUND READINGS)


#### Tab F  Mental Health over the Life Course: Later Life (BACKGROUND READINGS)


### Tab G  Mental Health Care and Policy (BACKGROUND READINGS)


### Tab H  General Guidelines for Breakout Session
This workshop will explore multiple levels of analysis, including environmental, sociocultural, behavioral, and biological, to see how these factors affect women’s mental health across the life course and across different racial/ethnic groups. The workshop will result in a rapporteur-prepared brief that summarizes the workshop presentations and discussions; the published brief will inform future research as well as program and policy discussions.

8:30 – 8:45 am  Introductions, Roadmap for the Workshop
Debra Umberson, University of Texas at Austin

8:45 – 9:00 am  Statement of Interest from Workshop Sponsor
Nicole Greene, Acting Director, Office on Women’s Health, U.S. Department of Health and Human Services
ADM Brett P. Giroir, Assistant Secretary for Health, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services

9:00 – 10:00 am  The Big Picture: Life Course Framing, Population Patterns, Measurement, Methods
Moderator: Jason Schnittker, University of Pennsylvania

- Conceptualizing and applying a sex/gender approach to women’s mental health
  Kristen Springer, Rutgers University

- Measurement advances toward understanding women’s mental health
  Nicholas Eaton, Stony Brook University
➢ **Vulnerable Populations: Differences Among Women Make A Difference**
  Vickie Mays, University of California, Los Angeles

10:00 – 10:15 am  BREAK

10:15 – 10:55 am  Psychological and Structural Factors in Women’s Mental Health  
{TAB C}

Moderator: Debra Umberson, University of Texas at Austin

➢ Psychology of gender  
  Vicki Helgeson, Carnegie Mellon University

➢ Structural factors that affect mental health over the life course  
  Deborah Carr, Boston University

10:55 – 11:35 am  Mental Health over the Life Course: Adolescence  
{TAB D}

Moderator: Bridget Goosby, University of Nebraska-Lincoln

➢ Life course epidemiology and adolescent mental health  
  Katherine Keyes, Columbia University [by videoconference]

➢ Mental health services and underserved youth populations  
  Victoria Ojeda, University of California, San Diego

11:35 am – 12:30 pm  LUNCH
12:30 – 1:10 pm  Mental Health over the Life Course: Mid-Life  

Moderator: Jennifer Payne, Johns Hopkins University

- Is it me or is it hot in here? Mood, memory, and Menopause  
  C. Neill Epperson, University of Pennsylvania

- A life course approach to understanding sex differences in depression and inflammation in mid-life  
  Natalie Slopen, University of Maryland College Park

1:10 – 1:50 pm  Mental Health over the Life Course: Later Life  

Moderator: Lisa Berkman, Harvard University

- What Do We Know (and Not Know) about the Gender Difference in Depression in Old Age?  
  Joan Girgus, Princeton University

- African American elderly women and mental health: Social work perspective  
  Karen Lincoln, University of Southern California [by videoconference]

1:50 – 2:00 pm  BREAK

2:00 – 3:15 pm  Mental Health Care and Policy  

Moderator: Hortensia Amaro, University of Southern California

- Assessment, identification diagnosis  
  Jeanne Miranda, University of California, Los Angeles

- Policies that constrain and promote women’s mental health  
  Chloe Bird, RAND Corporation
Gender-specific treatment
Christine Grella, University of California, Los Angeles

Work/family policy and women’s well-being
Jennifer Glass, University of Texas at Austin

3:15 – 3:30 pm  BREAK

3:30 – 4:15 pm  Breakout Groups, Brainstorming: Gaps, next steps, specific recommendations (moderators/steering committee members will facilitate breakout groups)  {TAB H}

4:15 – 5:00 pm  Group Reports: Agenda for Advancing Women’s Mental Health over the Life Course  {TAB H}
- What we need to know going forward
  o Gaps in data and knowledge
  o Most pressing challenges for women’s mental health
- What we need to do going forward (data, research, policy, practice)
  o Most promising and innovative directions

5:00 pm  Adjourn
GUIDANCE TO PRESENTERS

Throughout the sessions, attention to:

1. **Life course perspective, Etiology of Mood and Anxiety Disorders**
   - The intersecting biological, social, psychological, behavioral pathways → onset and trajectories of mood and anxiety disorders, with focus on gender/sex mental health disparities by race and ethnicity; **Recognition of early life origins** of biological, social, psychological risks/resources that contribute to lifelong patterns of women’s mental health

2. **Disparities and population patterns of mental health** – during specific periods of the life course and accumulating over the life course; intersectionality
   - Race, ethnicity, place, SES, sexual and religious minorities, immigration status, SES

3. **Multiple and intersecting levels of analysis**: Biological, psychological, social, behavioral risk and protective factors that may vary by race, ethnicity across the life course

4. **Moving forward**: Demographic trends, gaps in data and knowledge, key research questions

5. **Fostering resilience and identifying protective factors as well as attention to risk factors**

6. **Cutting-edge and forward-thinking data, methods**
**Workshop Steering Committee and Presenter Biosketches**

**Workshop Steering Committee Members**

**Debra J. Umberson (Chair)** is professor of sociology and director of the Population Research Center at the University of Texas at Austin. Professor Umberson's research focuses on social factors that influence population health with a particular emphasis on aging and life course change, marital and family ties, and gender and racial variation in health disparities. Her recent research, supported by a Robert Wood Johnson Foundation Investigator in Health Policy Research Award and the National Institute on Aging, examines how marital relationships affect health-related behavior and health care, and how those processes vary across gay, lesbian, and heterosexual unions. She is the past editor of the Journal of Health and Social Behavior and current chair of the Medical Sociology Section of the American Sociological Association. Dr. Umberson is an elected Fellow of the Gerontological Society of America, the 2015 recipient of the Matilda White Riley Distinguished Scholar Award from the American Sociological Association's Section on Aging and the Life Course and the 2016 recipient of the Leonard I. Pearlin Award for Distinguished Contributions to the Sociological Study of Mental Health from the American Sociological Association's Section on Mental Health. In her newest research, she focuses on black/white differences in exposure to the death of family members across the life course and the implications for long-term health and mortality disparities. She received her Ph.D. in Sociology from Vanderbilt University.

**Hortensia Amaro** is Dean's Professor of Social Work and Preventive Medicine, and associate vice provost of Community Research Initiatives at the University of Southern California. She has dramatically advanced the understanding of substance abuse disorder treatment, HIV prevention and other urgent public health challenges through a distinguished career that has spanned scholarly research, translation of science to practice, top-level policy consultation and service on four Institute of Medicine committees. Before joining USC in 2012, Dr. Amaro was with Northeastern University for 10 years, serving as associate dean and distinguished professor of health sciences and counseling psychology of the Bouvé College of Health Sciences, and as founder and director of the university's Institute on Urban Health Research. For 18 years prior to that, she was professor in the Boston University School of Public Health and in the Department of Pediatrics at the Boston University School of Medicine. She has received numerous awards, most recently the APHA Elizabeth Beckman Professors Who Inspire Award (2014) and the Sedgwick Memorial Medal for Public Health Service (2015). She has authored more than 140 scholarly publications, many widely-cited, and she has made landmark contributions to improving behavioral health care in community-based organizations by launching addiction treatment programs that have helped thousands of families and informing practice in agencies around the world. Dr. Amaro is a member of the National Academy of Medicine (2010) and currently serves on the Health and Medicine Division (HMD) Board on Population Health and Public Health Practice. In the past she has served on the the HMD Standing Committee on Integrating New Behavioral Health Measures in the Substance Abuse and Mental Health Services Administration's Data Collection Programs and the HMD Workshop Steering
Lisa F. Berkman is the director of the Harvard Center for Population and Development Studies (HCPDS) and the Thomas D. Cabot Professor of Public Policy, Epidemiology, and Global Health and Population at the Harvard T.H. Chan School of Public Health. She is an internationally recognized social epidemiologist whose work focuses extensively on social and policy influences on health outcomes. Her research orients toward understanding inequalities in health related to socioeconomic status, different racial and ethnic groups, and social networks, support and isolation. Dr. Berkman is the principal investigator of the Health and Aging Study in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI), a program project funded by the National Institute on Aging. HAALSI aims to study the drivers and consequences of HIV and non-communicable diseases in an aging population in Agincourt, South Africa. She is currently a member of the Conseil Scientifique de l’Institut de Recherche en Sante Publique (IReSP) in France and a member of NAM. She has been actively involved since 1994 on the GAZEL study, a cohort of 20,000 French employees of EDF-GDF, the large natural gas-electricity company. In May 2017, Dr. Berkman was appointed faculty director of the Ph.D. program in population health sciences at the Harvard T.H. Chan School of Public Health. She is author or co-author of several books and 300 publications/chapters. In 2003, she co-edited (with Ichiro Kawachi) Social Epidemiology, a groundbreaking textbook on this burgeoning field. A second edition was published in 2014 with co-authors Kawachi and Maria Glymour. In recent years, she served as a co-PI on the Work, Family & Health Network, a study involving workplace practices and employee and family health. From 2002-2016, she served as co-site director of the Robert Wood Johnson Foundation Health & Society Scholars program. Prior to becoming director of the HCPDS, Dr. Berkman was chair of the Department of Society, Human Development and Health at the Harvard T.H. Chan School of Public Health and was former head of the division of chronic disease epidemiology at Yale University. She received her Ph.D. in Epidemiology from University of California, Berkeley.

Bridget J. Goosby is currently a Happold Associate Professor of Sociology and director of the Biosociology of Minority Health Disparities Lab at the University of Nebraska-Lincoln. Upon completion of her Ph.D., she spent two years as a research analyst at the American Institutes for Research in Washington, D.C. From 2005-2007, she was a National Institute of Mental Health Postdoctoral Minority Mental Health Disparities Fellow at the University of Michigan in the Institute for Social Research’s Program for Research on Black Americans. Her primary research interest focuses on how the intersections of social inequality and discrimination ‘get under the skin’ to influence minority group members’ health over the life course and across generations. Physiologic stress response is a key pathway through which health disparities emerge in the U.S. population, thus her research integrates biological markers of stress, immune function, and chronic disease with innovative approaches to sociological data collections and analyses. To understand how these processes unfold to shape health risk and why these processes matter, her lab is currently implementing a variety of studies including several that examine how the stress of social exclusion and racial discrimination influence the health and well-being of parents and
their children in the local community. She earned her Ph.D. in Sociology and Demography from the Pennsylvania State University in 2003.

Jennifer L. Payne is associate professor of psychiatry and behavioral sciences, and director of the Women’s Mood Disorders Center, at Johns Hopkins University. Dr. Jennifer Payne’s research interests include clinical trials of novel therapeutics in depression and bipolar disorder; hormonal influences on mood and mood disorders; women and mood disorders and the genetics of depression. An author of over 40 publications, she presents widely on the above areas of interest most specifically focusing on women. Dr. Payne is a graduate of Washington University where she received her medical degree.

Jason Schnittker is a sociology professor at the University of Pennsylvania. His research interests are in the social, cultural, biological, and institutional determinants of health. His current research is concerned with the effects of incarceration on the health of individuals, families, and communities, funded in part by a Robert Wood Johnson Foundation Investigator Award in Health Policy. He is also interested in gene-environment relationships as they relate to health and well-being. He has a long-standing interest in public beliefs about mental health and illness, especially regarding how the public thinks about the causes of mental illness. He received his Ph.D. in Sociology from Indiana University.

Workshop Presenters

Chloe E. Bird is a senior social scientist at the RAND Corporation, where she studies women's health and determinants of gender differences in health and health care. She is also a member of the Pardee RAND Graduate School faculty. She recently served as editor-in-chief of the journal Women's Health Issues, where she is now associate editor. Her recent work includes a study assessing and mapping gender disparities in quality of care for cardiovascular disease and diabetes, among VA patients in California and Texas, and NIH-funded research on the impact of neighborhoods and behaviors on allostatic load, morbidity, and mortality. In her book Gender and Health: The Effects of Constrained Choice and Social Policies (Cambridge University Press, 2008), Bird and coauthor Patricia P. Rieker explore how policymakers and other stakeholders shape men's and women's opportunities to pursue a healthy life. They emphasize the need for research that informs stakeholders' decisions in order to improve women's health and reduce disparities. Bird is working to build a donor-funded Women's Heart Health Research and Policy Center at RAND to improve women's health by addressing deficits in women's health and health care. Bird received her Ph.D. in sociology from the University of Illinois at Urbana-Champaign.

Deborah Carr is Professor of Sociology at Boston University. She earned her Ph.D. in Sociology from the University of Wisconsin-Madison in 1997. Dr. Carr has held faculty positions at University of Michigan, University of Wisconsin, and most recently at Rutgers University, where she was acting director of the Institute for Health, Health Care Policy & Aging Research. Her research interests include aging and the life course, psychosocial factors influences on health over the life course, and end-of-life issues. She has published more than 100
refereed journal articles and book chapters focused on the ways that social roles, statuses, and transitions, including marriage, widowhood and divorce, work-family roles, caregiving, and body weight affect mental health in mid- and later-life. Her latest book project, *Golden Years?: Social Inequalities in Later Life* (under contract, Russell Sage) delves into the ways that persistent race, class, and gender inequalities shape experiences of old age in the United States. Carr is currently editor-in-chief of *Journal of Gerontology: Social Sciences* (2015-2020). She served as Chair of the Board of Overseers of the General Social Survey, and is a co-investigator on the Midlife Development in the United States (MIDUS) and Wisconsin Longitudinal Study (WLS). She was recently named principal investigator of the National Longitudinal Study of Youth 1979 (NLSY79), and is serving a four-year term (2015-19) on the National Institutes of Health’s Social Sciences and Population Studies B (SSPB) study section.

**Nicholas R. Eaton** is an Assistant Professor of Psychology at Stony Brook University, and he is Associated Faculty with the program in Women's, Gender, & Sexuality Studies. He completed his undergraduate education at Washington University in St. Louis and received his Ph.D. in Clinical Psychology from the University of Minnesota in 2012. His research involves mental disorder classification, dimensional conceptualizations of psychopathology, and understanding of how oppression (e.g., discrimination based on race/ethnicity, sexual orientation) relates to mental health outcomes. He has published around 80 peer-reviewed journal articles and book chapters and has received several awards, including the Robins-Guze Early Career Award from the American PsychoPathological Association and the Rising Star designation from the Association for Psychological Science.

**C. Neill Epperson** is Director of the Penn Center for Women's Behavioral Wellness and Professor (with Tenure) of Psychiatry and Obstetrics/Gynecology at the Perelman School of Medicine at the University of Pennsylvania where she is also the Director of the Penn Building Interdisciplinary Research Careers in Women’s Health Program and Penn PROMOTES Research on Sex and Gender in Health. Dr. Epperson was recruited to the University of Pennsylvania from Yale University School of Medicine in the Fall of 2009 to initiate and direct the department’s division of women’s behavioral health. The overarching goal of Dr. Epperson's research has been the neuroendocrinology of mood, behavior, and cognition across the female life span. To this end, her laboratory utilizes a wide range of scientific tools, including brain imaging, psychophysiology, hormone treatment, amino acid depletion and pharmacologic challenge. She has published more than 190 peer-reviewed manuscripts, chapters, reviews, and meeting abstracts. Dr. Epperson’s research is funded by the National Institute of Mental Health, that National Institute on Drug Abuse, the National Institute on Aging, the National Cancer Institute, and the Office of Research on Women’s Health. She received her M.D. from the University of North Carolina at Chapel Hill.

**Joan Girgus** is a Professor of Psychology Emeritus at Princeton University. She retired on July 1, 2017 after 40 years of teaching and research as a Professor of Psychology. During those 40 years, she also served as Dean of the College, Chair of the Psychology Department, and Special Assistant to the Dean of the Faculty for diversity issues with an emphasis on women faculty and making Princeton a more family-friendly university. Prior to coming to Princeton, she served as a faculty member and dean at the City College of the City University of New York. Dr. Girgus has done research and written books and papers on perception and perceptual development, the transition from childhood to adolescence, the psychosocial bases of depression, gender
differences in depression, and sensitivity to interpersonal feedback. She has also written papers on undergraduate science education, on women in science, and on making universities more family-friendly organizations. Dr. Girgus was one of the principals of The Learning Alliance, the first just-in-time provider of strategic expertise to college and university leaders. From 1987-99, she directed the Pew Science Program, a national program to improve undergraduate science education sponsored by the Pew Charitable Trusts. Dr. Girgus received her B.A. from Sarah Lawrence College and both her M.A. and Ph.D. from the Graduate Faculty of the New School for Social Research in New York City. She has served as a trustee of Adelphi University, the American Association of Higher Education (AAHE), McCarter Theatre, the Princeton Adult School, the Princeton Senior Resource Center, Sarah Lawrence College and the Wenner-Gren Foundation.

**Jennifer Glass** is the Centennial Professor of Liberal Arts in the Department of Sociology and the Population Research Center of the University of Texas, Austin. She has published over 60 articles and books on work and family issues, gender stratification in the labor force, mother’s employment and mental health, and religious conservatism and women’s economic attainment, with funding from the National Science Foundation, the National Institutes of Health, and the Alfred P. Sloan Foundation. She received the Reuben Hill Award from the National Council on Family Relations, Best Paper Award from the Family Section of the American Sociological Association, and thrice been nominated for the Rosabeth Moss Kanter Award for Excellence in Work-Family Research. She has chaired the Sex and Gender Section, the Family Section, the Organizations and Work Section, and served as Vice-President of the American Sociological Association. She is currently the Executive Director of the Council on Contemporary Families, and Chair of the Social Sciences and Population Study Section at the National Institutes of Health. Her most recent projects explore whether governmental work-family policies improve parents’ well-being, whether women’s jobs really have better work-family benefits than men’s, why women's retention in STEM occupations remains so abysmally low, and how the economic costs of motherhood have changed over time, all as part of a larger project to understand the roots of women’s disadvantage in the labor market. She received her Ph.D. in sociology from the University of Wisconsin, Madison.

**Christine E. Grella** is Professor-in-Residence in the Department of Psychiatry and Biobehavioral Sciences at the University of California, Los Angeles and Co-Director of the UCLA Integrated Substance Abuse Programs (ISAP). She received her doctorate in Social Psychology from the University of California, Santa Cruz in 1985. Dr. Grella has over 30 years of experience conducting research on a wide range of prevention and intervention projects on substance use among youth and adults. Her research has focused on service utilization and outcomes of participation in multiple service delivery systems, including substance use disorder treatment, mental health, child welfare, health services, HIV services, and criminal justice. She has led several longitudinal cohort studies, conducted over a dozen evaluation studies of treatment interventions for individuals with substance use and/or mental health disorders, and conducted epidemiological studies using national survey data. She has published over 40 peer-reviewed articles that examine various aspects of the influence of gender and gender differences in substance use disorder treatment for women. Her research in this area has been instrumental in documenting the development of gender-specific or “gender-sensitive” treatment interventions that incorporate a focus on women’s specific therapeutic needs within a comprehensive
framework. The findings from her research have been widely published in the areas of addiction, mental health, health services, and evaluation research. Dr. Grella also serves as a reviewer for NIH/NIDA and on several advisory boards for community organizations and foundations.

**Vicki S. Helgeson** is Professor of Psychology at Carnegie Mellon University. She has studied how people adjust to chronic illness for 30 years, including people with heart disease, breast cancer, prostate cancer, and, most recently, diabetes. She has received steady funding from NIH since 1993 to support her work. Her research focuses on the implications of both person factors (e.g., perceived control, gender-related traits) as well as social environment factors (e.g., family relations, peer relations) for psychological and physical health. Dr. Helgeson has utilized a variety of designs in her work, including experimental, ecological momentary assessment, and interventions. Notably, she completed a 10-year longitudinal study of women with breast cancer and a 15-year longitudinal study of youth with type 1 diabetes. She is a fellow of the Society of Behavioral Medicine as well as the Society of Personality and Social Psychology and has served as Associate Editor for *Personality and Social Psychology Bulletin*, *Psychological Science*, and the *Journal of Consulting and Clinical Psychology*, and *Women’s Health: Research on Gender, Behavior, and Policy*. Dr. Helgeson served on the NIH Psychosocial Risk and Disease Prevention Study Section and has served as an ad hoc reviewer on a number of special emphasis and fellowship panels. She received her Ph.D. in social psychology from the University of Denver.

**Karen D. Lincoln** is an Associate Professor in the Suzanne Dworak-Peck School of Social Work, Director of the USC Hartford Center of Excellence in Geriatric Social Work, Co-director of the Southern California Clinical and Translational Science Institute – Community Engagement Core, Senior Scientist at the Edward R. Roybal Institute on Aging, and Founder and Director of Advocates for African American Elders at the University of Southern California. Dr. Lincoln has published over 60 articles and book chapters in the areas of stress, aging, and mental health disparities. She has received more than $2.8 million in grant funding to support her research which focuses on improving clinical and community-based treatment of African Americans with mental health disorders and chronic health conditions. Dr. Lincoln is a Fellow of the Gerontological Society of America and a Hartford Faculty Scholar. Dr. Lincoln is an honors graduate from UC Berkeley where she received a B.A. in Sociology with a minor in African American studies and a graduate from the University of Michigan where she earned a MSW, a M.A. in Sociology and a Ph.D. in Social Work and Sociology.

**Vickie Mays** is a Professor in the Department of Psychology in the College of Letters and Sciences, as well as a Professor in the Department of Health Services. Professor Mays is also the Director of the UCLA Center on Research, Education, Training and Strategic Communication on Minority Health Disparities. She teaches courses on health status and health behaviors of racial and ethnic minority groups, research ethics in biomedical and behavioral research in racial/ethnic minority populations, research methods in minority research, as well as courses on social determinants of mental disorders and psychopathology. She holds a Ph.D. in Clinical Psychology and an M.S.P.H. in Health Services, with postdoctoral training in psychiatric epidemiology, survey research as it applies to ethnic minorities (University of Michigan) and health policy (RAND). Professor Mays' research primarily focuses on the mental and physical health disparities affecting racial and ethnic minority populations. She has a long history of research
Jeanne Miranda is a Professor in the Department of Psychiatry and Biobehavioral Sciences at UCLA. Her major research contributions have been in evaluating mental health care for poor and minority women. She is evaluating an intervention her team developed for families adopting older children from foster care, adapting depression interventions for young women in Uganda, as well as evaluating governmental social programs there. She has developed and is evaluating a resilience intervention for low-income and minority LGBT populations. She was the Senior Scientific Editor of Mental Health: Culture, Race and Ethnicity: A Report of the Surgeon General. She became a member of the Institute of Medicine, now the National Academy of Medicine, in 2005. Dr. Miranda is the 2008 recipient of the Emily Mumford Award for Contributions to Social Medicine from Columbia University. She received her PhD in Clinical Psychology from the University of Kansas.

Victoria D. Ojeda, is an Associate Professor, in the Division of Global Public Health in the Department of Medicine at University of California, San Diego School of Medicine. Her current research focuses on the health of underserved and vulnerable populations, including justice-involved persons and persons living with serious mental illness, particularly transition age youth (ages 18-26 years). She has published on health services issues (e.g., access to health insurance coverage, utilization of health and mental health services), HIV prevention, and migrant health issues. Dr. Ojeda is the Principal Investigator of the UCSD RE-LINK program, funded by the DHHS-Office of Minority Health (grant #: CPIMP161137), which aims to reduce disparities in health status and access to care among young adults ages 18-26 who are exiting the justice system and seeking to reenter the community. Dr. Ojeda is also PI of the NIH-National Institute of Minority Health and Health Disparities Grant R01 MD011528: Impact of Peer Support on Health Service and Social Disparities Among Minority Youth with SMI; this mixed methods study aims to assess whether peer-support reduces disparities in mental health service utilization and social outcomes among transitional age youth (TAY) age 16-24 with severe mental illness (SMI). Dr. Ojeda is also Co-investigator of studies examining mental health service delivery in San Diego and Orange Counties following implementation of the California Mental Health
Services Act. In addition, Dr. Ojeda is the founder of the UCSD Clean Slate Free Tattoo Removal Program (funder: San Diego Indigent Criminal Legal Defense Fund), which offers free laser tattoo removal services to adults on Probation in San Diego County; and she is Co-Director of the Health Frontiers in Tijuana Free Clinic, which receives support from the Elton John AIDS Foundation. Dr. Ojeda holds a BA in Psychology and Spanish from Brandeis University. She obtained her MPH and PhD in Public Health (emphasis: Community Health Sciences) from the School of Public Health at UCLA. Dr. Ojeda completed a NIMH Post-doctoral fellowship in Mental Health Policy at the Department of Health Care Policy at the Harvard Medical School.

Natalie Slopen is an Assistant Professor in the Department of Epidemiology and Biostatistics at the University of Maryland College Park. Dr. Slopen’s research focuses on social influences on health, health disparities, and mechanisms through which childhood experiences are embedded to increase risk for later chronic diseases. The overarching goal of her research is to identify processes and conditions that can be targeted by interventions in order to reduce health disparities and promote health over the life course. Dr. Slopen completed her Doctorate of Science in Social Epidemiology at the Harvard School of Public Health, and her postdoctoral fellowship training at the Center on the Developing Child at Harvard University.

Kristen W. Springer is Associate Professor of Sociology and an Affiliate of the Institute for Health, Health Care Policy, and Aging Research at Rutgers University. Her research centers on gender and health inequalities, with a focus on the examining the interactive influence of biology and social environment on men’s physical health. Her recent research explores on how masculinity leads to men’s poorer health through avoiding doctor visits and through increased cardiovascular and neuroendocrine stress reactions. She has published in leading health and gender journals, has co-edited a Special Issue of Social Science and Medicine on gender and health, and has been featured in top international and U.S. new outlets. She received her Ph.D. in sociology at the University of Wisconsin-Madison.
The National Academies of 
SCIENCES • ENGINEERING • MEDICINE

Women’s Mental Health across the Life Course through a Sex-Gender Lens: A National Academy of Sciences, Engineering, and Medicine Workshop

PURPOSE

The purpose of this order is to support a one-day workshop on the Women’s Mental Health Across the Life Course Through a Sex-Gender Lens. This workshop will explore the multiple levels of analyses, which include environmental, sociocultural, behavioral, and biological, to see how these factors impact women’s mental health at different stages of life and across different racial/ethnic groups.

To that end, the DHHS, Office of Women’s Health (OWH) wants to bring together scientific and public health experts who have used multi-level analytic approaches in mental health research to report outcomes by age, sex, and race/ethnicity. The workshop audience should include key stakeholders and experts from across clinical, behavioral, and public health disciplines in the public and private sector. Workshop panel member presentations should answer the following questions:

- What are the useful multilevel (i.e., environmental, sociocultural, behavioral, and biological) methodologies to understand mental health issues in adolescent girls, young adults, adults, and older women in clinical and public health research; and reduce racial/ethnic disparities in women’s mental health?
- What are the research gaps in the body of multilevel research of mental health in women?
- What types of multilevel and/or multi-disciplinary interventions, programs, and policies are needed to address sex-gender differences and health disparities in mental health?

Expected outcomes from this workshop include:

- Published workshop summary report to inform future program and policy recommendations for OWH and its partners.
- A live and archived webcast of the workshop.
- Nationwide dissemination plan for the workshop summary report, which will guide important future directions for research on women's mental health.

BACKGROUND

The Institute of Medicine (IOM), at the National Academies of Sciences, Engineering, and Medicine (“National Academies”), concluded in a report that insufficient attention has been paid to improving disease-related morbidity, wellness, and quality of life as health outcomes for women and a sub-category or racial/ethnic women (IOM, 2010). Sex and gender have typically not been a priority for the field of health disparities. In September 2015, the Committee on
Population and the Board on Population Health and Public Health Practice of the National Academies convened a workshop, sponsored by Office of Research on Women’s Health (ORWH), of the National Institutes of Health (NIH), on the important determinants, consequences, effects, and issues attending the comparative deficiency in the health of U.S. women in comparison to other economically advanced countries. ORWH recently released the 4th Edition of the Women of Color Health Databook, which highlights the different patterns of health disparities and health determinants based on the stratification of women by race/ethnicity. As of January 2016, NIH-funded scientists are required to account for the possible role of sex as a biological variable in vertebrate animal and human studies. The current NIH Health Disparities Strategic Plan represents the major priorities and broad range of activities that the NIH ICs will undertake towards ultimately eliminating health disparities. One of the three major goals the plan outlines is that each NIH Institute and Center must strive to achieve is to conduct and support intensive research on the factors underlying health disparities.

Increased knowledge is needed about the factors relevant to gender health disparities and the life course for women. In order to better define causal pathways to health and disease, the identification and the linkage of environmental, sociocultural, behavioral, and biological factors must occur (Hill et al, 2015. Ethnicity & Disease). NIH’s National Institute on Aging (NIA) supports health disparities research based upon a health disparities research framework which includes these multiple levels of analyses. Developing rigorous research methodologies for studying women’s health issues over the life course is critical and has not been a primary focus of public health research. The life course framework has been applied and promoted often for understanding maternal child health and reproductive health issues. The World Health Organization (WHO) recently released The World Report on Ageing and Health, which outlines a life course approach for action to foster healthy ageing built around the new concept of functional ability.

The U.S. Department of Health and Human Services (HHS), Office on Women’s Health (OWH) has a mission to provide national leadership and coordination to improve the health of women and girls through policy, education and model programs. Through grants and contracts, OWH funds a wide spectrum of activities and programs in support of its mission. OWH works with many partners, including federal government agencies, nonprofit organizations, consumer groups, associations of health care professionals, tribal organizations, and state, county, and local governments. OWH has several focus areas, two of which are on health across the lifespan and health disparities.

Accordingly, OWH wants to sponsor a one-day workshop, similar to the one the Committee on Population and the Board on Population Health and Public Health Practice of the National Academies convened for ORWH, that will be organized to support the development of scientific knowledge and foster a critical understanding around the multilevel approaches to

1 http://orwh.od.nih.gov/resources/policyreports/pdf/WoC-Databook-FINAL.pdf
2 NIH Guide Notice NOTOD-15-002
researching women’s health issues across the life course to improve the mental health of diverse populations of women.

WORKSHOP STEERING COMMITTEE AND PRODUCTS

After discussions with staff members from the Office on Women’s Health (OWH), the Committee on Population will nominate a steering committee to plan and oversee the workshop and planned dissemination activities; members of the steering committee will have expertise in research on gender, multi-level and life course analyses, women’s mental health, and health disparities. Working with CPOP and OWH staff members, the steering committee will specify an agenda, identify participants and presentation topics, and convene a public workshop and disseminate workshop summary brief. The products of the workshop will be: 1) an unedited transcript of the meeting; 2) electronic versions of the agenda, presenter biosketches, and all input from the presenters, including presentation materials and slides from the workshop; 3) a published workshop summary brief to shape future research and to inform program and policy recommendations for OWH and its partners; 4) a live and archived webcast of the workshop to maximize participation; and 5) dissemination of the workshop summary brief, including presentations on the workshop discussions at one or more regional or national scientific conferences or meetings.

A rapporteur-prepared brief of approximately 20 manuscript pages that summarizes the workshop presentations and discussions will increase awareness in the population and health research community of the causes and consequences of sex-gender differences and disparities in mental health across the life course. The brief would be organized around the workshop questions and outcomes outlined above. The published brief (known as a “proceedings of a workshop in brief”) would inform future research as well as program and policy discussions.

The brief would be prepared the month after the workshop and would be reviewed by a small number of experts, including one or two who attended the workshop and one or two who did not in order to ensure overall clarity. Needed revisions would be completed, and the brief approved by the Academies. Dissemination of the brief will include publication online and distribution to audiences identified by the OWH, CPOP, and the steering committee as well as distribution to all workshop participants.
Committee on Population (CPOP)

Roster

Kathleen Mullan Harris (Chair)
James E. Haar Distinguished Professor,
Sociology Department
University of North Carolina at Chapel Hill

Mark D. Hayward (Alternate Chair)
Professor of Sociology
Centennial Commission Professor in the
Liberal Arts
Research Associate, Population Research Center
University of Texas at Austin

Jere R. Behrman
W.R. Kenan Jr. Professor of Economics and Sociology
Department of Economics
University of Pennsylvania

Lisa F. Berkman (NAM)
Thomas Cabot Professor of Public Policy
Epidemiology and Population and International Health
Harvard School of Public Health

Ann K. Blanc
Vice President
Social and Behavioral Science Research
Population Council

Vicki A. Freedman
Research Professor
Institute for Social Research
University of Michigan

Dana A. Glei
Senior Research Investigator
Georgetown University

Hillard S. Kaplan
Professor
Department of Anthropology
University of New Mexico

Sara S. McLanahan
William S. Todd Professor of Sociology and Public Affairs
Princeton University

Emilio A. Parrado
Professor and Chair
Department of Sociology and Population Studies Center
University of Pennsylvania

David R. Weir
Research Professor, Survey Research Center
Institute for Social Research
University of Michigan

John R. Wilmoth
Director, Population Division/DESA
United Nations

STAFF

Malay K. Majmundar, Director

Mary Ghitelman, Senior Program Assistant
Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health

Kristen W. Springer a,*, Jeanne Mager Stellman b, Rebecca M. Jordan-Young c

a Department of Sociology and Institute for Health, Health Care Policy and Aging Research, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901, United States

b Department of Health Policy and Management, Mailman School of Public Health, Columbia University, United States

c Department of Women's, Gender, and Sexuality Studies, Barnard College, Columbia University, United States

ARTICLE INFO

Article history:
Available online 15 June 2011

Keywords:
Sex differences
Gender differences
Gender and health
Health
Institute of medicine report
Biosocial
Sex/gender

ABSTRACT

Extensive medical, public health, and social science research have focused on cataloguing male–female differences in human health. Unfortunately, much of this research unscientifically and unquestionably attributes these differences to biological causes — as exemplified in the Institute of Medicine’s conclusion that “every cell has a sex.” In this manuscript we theorize the entanglement of sex and gender in human health research and articulate good practice guidelines for assessing the role of biological processes — along with social and biosocial processes — in the production of non-reproductive health differences between and among men and women. There are two basic tenets underlying this project. The first is that sex itself is not a biological mechanism and the second is that “sex” and “gender” are entangled, and analyses should proceed by assuming that measures of sex are not pristine, but include effects of gender. Building from these tenets — and using cardiovascular disease as a consistent example — we articulate a process that scientists and researchers can use to seriously and systematically assess the role of biology and social environment in the production of health among men and women. We hope that this intervention will be one further step toward understanding the complexity and nuance of health outcomes, and that this increased knowledge can be used to improve human health.

© 2011 Elsevier Ltd. All rights reserved.

Introduction

Social science and health research ubiquitously treat sex as a dichotomized independent variable. The United States Institute of Medicine (IOM) Committee on Understanding the Biology of Sex and Gender Differences drew an ‘overarching conclusion’ that “Sex … should be considered when designing and analyzing studies in all areas and at all levels of biomedical and health-related research. …” (xix) and that “every cell has a sex” (Wizemann & Pardue, 2001, p.4; see also pp.28–44). Indeed, the IOM not only reified sex as a master independent variable, but also focused its own committee work on “sex-based differences, versus similarities, as …more likely to … lead to greater understanding of the significance of sex in human biology and health” (2). The IOM study did acknowledge that social and cultural factors could contribute to observed male-female differences in anatomy and physiology (see especially pp.18–19). Generally, though, the committee’s analysis departed from its official openness to such interactive models, and from its designation of “gender” as a continuum (p.17). Instead, as two dissenting committee members noted, the committee’s overall approach suggests a “predominance of physiology, with a subsequent fine-tuning by environment” (p.18), and moreover, as we argue below, such an approach incorrectly implies that scientists can completely “control” for gender to identify the pure substrate of sex biology that underlies observed male-female differences.

The IOM, drawing especially on definitions advanced by the World Health Organization (WHO) and the style manual of the Journal of the American Medical Association (JAMA), defines sex as “The classification of living things, generally as male or female according to their reproductive organs and functions assigned by chromosomal complement” and gender as “A person’s self-representation as male or female, or how that person is responded to by social institutions based on the individual’s gender presentation. Gender is rooted in biology and shaped by environment and experience” (Wizemann & Pardue, 2001 p.17). These definitions lend a superficial sense that sex and gender are distinct domains, even as they give causal and temporal priority to biology (“gender is rooted in biology” but sex is presumably pristine and emerges regardless of environment and experience).
Entanglement of sex/gender in human health

In this paper we build on prior research to provide an alternative to IOM-type mandates for health research where “sex” is treated as the main explanatory variable for proposed male—female health differences in human health (Bird & Rieker, 1999; Epstein, 2007; Fausto-Sterling, 2005). We aim to articulate a theoretical frame and good practice guidelines for exploring the role of sex-related biology in human health, especially as it may influence non-reproductive differences between males and females. We are explicitly engaging with sex difference research — so we take a frame of male—female difference in much of our discussion, although we also focus on the importance of within-sex variation. There are two basic tenets to our argument. First, sex is not a biological mechanism and its use as a proxy for other measures does little to further the understanding of health-related research questions. Second, in the vast majority of health research, “sex” and “gender” are entangled and analyses should proceed by assuming that measures of sex are not pristine, but include effects of gender. We are not arguing against any biologically based male—female differences. Rather we are arguing that the vast majority of male—female health differences are due to the effects of the irreducibly entangled phenomenon of “sex/gender,” and therefore this entanglement should be theorized, modeled, and assumed until proven otherwise.

We borrow the concept of entanglement from multiple other fields including quantum physics and science/technology studies, where entanglement has been used to describe inextricably interwoven factors (Callon & Rabeharisoa, 2003; Papp et al., 2009). For example, according to quantum physics, no individual element can fully describe a quantum mechanical state without incorporation of other elements, and any attempt to measure a single factor inevitably perturbs the remaining system rendering the results inaccurate. Of course, physicists do measure physical properties, and do accept a certain level of inaccuracy and imprecision in their work by doing so — just as social scientists define and measure sex and gender in research on health outcomes. Nonetheless, we maintain that a true and full appreciation of sex/gender in human health cannot rely on the dichotomization of people into males and females without attention to how sex and gender may be — and almost always are — entangled.

The idea of sex/gender entanglement is not new and a growing body of research demonstrates the problematic task of distinguishing between sex and gender in practice (Fausto-Sterling, 2005; Kessler, 1998; Oudshoorn, 1994). Feminist epidemiologists and biologists, among other scientists, increasingly replace the discrete concepts of “sex” and “gender” with more complex formulations, such as Nancy Krieger’s notions of “biologic expressions of gender” and “gendered expressions of biology” (Krieger, 2003, p. 653). As Kaiser, Haller, Schmitz, and Nitsch (2009) recently observed, “sex is not a pure bodily and material fact, but is deeply interwoven with social and cultural constructions of gender” (Kaiser et al., 2009, p.50). Fausto-Sterling (2005) has shown in the case of bone development and change, and Jordan-Young and Rumiti (forthcoming) have argued in the case of brain structure and function, that gendered life experiences have material effects on the body. These effects show up, in turn, as biologically based “sex differences.” Indeed, the social structure of gender even creates biological differences for boys and girls starting in the womb (e.g. birth weight) — as evidenced by lower levels of prenatal care for female fetuses in countries with a preference for boys (Al-Akour, 2008).

While the causal link between “sex” and “gender” is often thought to flow automatically from biological to social difference, recent research has forcefully demonstrated that the influence often operates in the other direction. A strong example can be found in social neuroendocrinology where research has confirmed that mood states, social interactions, and status differentials affect neuroendocrine production and function (Booth, Granger, Mazur, & Kivlighan, 2006; Haneishi et al., 2007; Sapolsky, 1997; van Anders & Watson, 2006). One obvious conclusion is that sex-linked outcomes related to neuroendocrine function must be conceived as flowing, at least in part, from social variables.

The potential implications of recognizing “sex endocrinology” as a biosocial complex are profound. Consider the relationship of neuroendocrine factors to risks for cardiovascular disease (CVD), which we will continue to use as an example throughout because of its epidemiological importance as a primary cause of death and because of the historical importance of CVD research originally ignoring women and then reacting by automatically focusing on male-female differences in sex-linked biology (Epstein, 2007; Mosca et al., 2004).

A multidisciplinary team working on psychosocial factors in CVD has noted, for example, that “psychosocial stress reliably induces ovarian dysfunction, hypercortisolemia, and excessive adrenergic activation in pre-menopausal females, leading to accelerated atherosclerosis” (Rozanski, Blumenthal, & Kaplan, 1999, p.2192). These results of psychosocial stress constitute variables gendered, and therefore characterized as aspects of “sex biology.” However, since psychosocial stress is itself the result of processes that are profoundly gendered (e.g., uneven family obligations, gender-specific forms of harassment or discrimination, and greater levels of poverty and wage insecurity), these “biological” factors are already inextricably bound up with the social world of gender (Rozanski et al., 1999). Despite some progress, we still know relatively little about specific social-biological pathways through which gendered arrangements become embodied as within- and between-sex differences. We therefore propose that the general dictum should be that sex and gender are entangled, rather than the opposite dictum as set forth by the IOM.

The difficulty in identifying the true directionality of the “causal” links leads us to follow Kaiser et al. (2009), and use the term sex/gender “to highlight this socio-biological intertwinement whenever possible” (Kaiser et al., 2009, p.50). This view of sex/gender fits very well with the contemporary focus on intersectional approaches to health whereby aspects of social status (e.g., gender, race, socioeconomic status, and sexuality) are understood to affect health outcomes in complex, multiplicative ways that can never properly be captured by attempts to parcel out the individual contributions of single social domains (Hankivsky & Christoffersen, 2008; Jackson & Williams, 2006).

We approach sex/gender in this intersectional light and conceptualize sex/gender as a domain of complex phenomena that are simultaneously biological and social, rather than a domain in which the social and biological “overlap.” We use the term gender (read: gender-not-sex) to draw attention to specifically social and structural factors, such as patterns in the distribution of family responsibilities, formal and informal sanctions for gender non-conforming behavior, and so on. (For recent, in-depth analyses that support the claim that these patterns cannot be attributed to “original” differences between males and females, see Fine, 2010; Jordan-Young, 2010; Jordan-Young & Rumiti, forthcoming). We avoid the use of the term sex (read: “sex-not-gender”) as a standalone indicator of biology, or as a broad reference to males and females (as in “sex differences”) because it is rarely specific enough to guide particular investigations, and because it is too easily confused with the more accurate composite phenomenon of sex/gender that we use here.

Further, the direction of influence should not be assumed to flow from an idealized version of “pure” biology outward to the behavioral and social world. Consider, for example, glucose metabolism,
Male–female differences in glucose metabolism have been reported (Basu et al., 2006; Peterson et al., 2007), and cellular level mechanisms related to glucose–insulin interactions have been hypothesized to explain this difference (Basu et al., 2006). Yet a recent experimental study of response to oral glucose “challenge” found that the male–female difference, comparable in size in this study to other reports, could be entirely explained by difference in skeletal muscle mass between men and women (Rattarasarn, Leelawattana, & Soonthornpun, 2010). Given the extensive data on social influences on the male–female difference in muscle mass (for instance, via differential nutrition, aesthetic norms, occupation, and use of outdoor spaces for exercise), the metabolic differences seen at the cellular level between males and females could be understood as bearing the traces of their gendered lives (Bentley et al., 1999; Courtenay, 2000; Richardson & Mitchell, 2010).

In a case such as this, sex/gender is a preferable notation to “sex” alone, although the exact qualities that sex/gender represent may not be fully specified. Unmeasured aspects of gender will always be present and usually we will be unaware of the specific ways in which gendering of activities, as through nutrition and psychology, affects cellular level processes.

**Good practice guidelines for examining male–female human health differences**

We build on the theoretical frame of sex/gender entanglement to present an encapsulation of good practice guidelines for researching male–female health differences (Fig. 1). This encapsulation is critical because these good research practices are often ignored in sex/gender and health research, which hampers the ability of research to provide fruitful insights into health inequalities and disparities. These guidelines are organized into three sequential sections (research plan development; *a priori* comparison plan; and interpretation and presentation of results), but as with all research, looping back through the sections is inevitable and desirable.

**Research plan development**

Because we posit that sex and gender are entangled, the first step is to assess whether or not biological male–female difference research is both warranted and possible. If yes, the next step is to *a priori* develop and articulate specific, testable hypotheses, with precise well-defined measures observable in both men and women, and to articulate plausible biological/biosocial mechanisms to explain the biological contribution to male–female health differences. (*Biological mechanism* is not meant to suggest that biology can be separated from the social environment, but rather to refer to specific mechanisms that can be described in molecular or biochemical terms.)

**Fully define outcome and independent measures *a priori***

An accurate understanding of the biological contributions of male–female differences requires precisely identifying and

<table>
<thead>
<tr>
<th>RESEARCH PLAN DEVELOPMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fully define outcome and independent measures <em>a priori</em></td>
</tr>
<tr>
<td>2. Identify plausible biological mechanism(s) explaining male–female difference</td>
</tr>
<tr>
<td>3. Directly test the biological mechanism(s) – if at all possible</td>
</tr>
<tr>
<td>4. Develop, operationalize, and test alternative hypotheses that include social and biosocial pathways.</td>
</tr>
<tr>
<td>5. Incorporate and test the biological and social/biosocial mechanisms together.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A PRIORI COMPARISON PLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Include appropriate control groups for ALL hypotheses (biological, social, biosocial) and accurately test for differences</td>
</tr>
<tr>
<td>2. Include sufficient measures for confounders and covariates</td>
</tr>
<tr>
<td>3. Attend to issues of statistical significance and sample size</td>
</tr>
<tr>
<td>4. Conduct sensitivity analyses</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERPRETATION AND PRESENTATION OF RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Present within group and between group differences</td>
</tr>
<tr>
<td>2. Explain effect size: substantive vs. significant difference</td>
</tr>
<tr>
<td>3. Report post-hoc tests and adjust significance levels appropriately</td>
</tr>
<tr>
<td>4. Describe social/biosocial attenuation <em>and</em> the purely social/biosocial findings</td>
</tr>
<tr>
<td>5. Describe similarities and differences for biological causes</td>
</tr>
<tr>
<td>6. Thoroughly report limitations of the study and provide enough information for replication</td>
</tr>
</tbody>
</table>

Fig. 1. Good practice guidelines for researching sex/gender in human health.
operationalizing the outcome of interest. Continuing with the example of CVD, we see that the rubric “cardiovascular disease” itself may be too broad for elucidating observed male-female differences in rates. The International Classification of Diseases (ICD-9) uses the codes 350 to 459 to classify circulatory system diseases. “Ischemic heart disease” uses more than 20 different codes. The 16th edition of Harrison’s Manual of Medicine includes 13 sub-sections on CVD (Kasper, Braunwald, Fauci, Hauser, Longo, & Jameson, 2005). True biological male—female differences in CVD would be expected to vary in important ways depending on the specific aspect of CVD under investigation.

CVD studies may also include a wide range of clinical signs and symptoms like chest or back pain (Berg, Bjorck, Dudas, Lappas, & Rosengren, 2009; Hendrix, Mayhan, Lackland, & Egan, 2005), serum progesterone level (Nilsson, Fransson, & Brismar, 2009), or a psychosocial risk factor like depression (Rozanski et al., 1999). One complication in defining measures is that physicians may differently diagnose and treat the same presentation of symptoms depending on whether the patient is male or female (Enríquez, Pratap, Zblit, Calvin, & Volgman, 2008).

In addition, studies of the “same” outcome can produce dramatically different results depending on the definitions and measures used to operationalize the outcome. For example, research on male-female differences in the CVD-related outcome of metabolic syndrome shows great variation in the size, shape, and even the presence of male-female differences depending on the definition of the outcome used in the study. As articulated by Regitz-Zagrosek et al. (2006): “Classification according to WHO criteria generally led to a 50% higher estimation of prevalence compared with the EGIR (European Group for the Study of Insulin Resistance) criteria in men. This can mostly be explained by the different cut-off value for central obesity used in the WHO definition. For women, the difference was smaller.” (p.137).

Sex/gender differences in a specifically defined outcome may be partially explained by other factors (Naqvi, Naqvi, & Merz, 2005; Rozanski et al., 1999; Suarez, 2006). Duprez et al. (2009) found significant male-female differences in arterial elasticity — an indicator of progression to cardiovascular morbidity and mortality — that were completely accounted for by height differences between men and women. Importantly, the broad variability in height within both sexes means that using sex as a proxy for arterial elasticity results in, at best, unnecessary loss of information and, at worst, gross misclassification errors leading to erroneous results.

Identify plausible biological mechanism(s) explaining male—female difference

Studies should be designed to explicate plausible biological mechanisms and “sex,” as such, is not a mechanism. Putative male-female difference must be tied to plausible biological mechanisms and operationalized into measurable variables. A thorough, critical investigation of the literature will assist in understanding biological mechanisms and how they can be operationalized before proceeding with hypothesis development. How the mechanism(s) would be different in men and women — on average — must be articulated. For example, male-female differences in mortality have been linked to male-female variations in metabolic syndrome and diabetes, a CVD risk factor. As explained by Regitz-Zagrosek, Lehmkuhl, and Weickert (2006): “Lipid accumulation patterns differ between women and men. Pre-menopausal women more frequently develop peripheral obesity with subcutaneous fat accumulation, whereas men and post-menopausal women are more prone to central or android obesity. In particular, android obesity is associated with increased cardiovascular mortality and the development of type 2 diabetes” (p.136). Here the proposed biological mechanism for male-female difference is not for all females and males — but for pre-menopausal women compared to men (and post-menopausal women), underscoring the complexity (and generally unacknowledged variability) of the categories “male” and “female.” (It is also possible that differential deposition of fat is socially created, as discussed below.)

Directly test the biological mechanism(s) — if at all possible

Identifying a male—female difference in an outcome does not tell us anything about the possible mechanisms (biological or social) that may produce this difference. Therefore, using sex as a variable is an inadequate proxy for testing a proposed biological mechanism. For example, if the proposed biological pathway for CVD mortality is central or android obesity, then it is more useful to directly measure men’s and women’s obesity, rather than simply including a “sex” variable and then hypothesizing that the effect is due to a difference (on average) in the location of men’s and women’s fat stores. Even multilevel dichotomous measures, like male, pre-menopausal female, and post-menopausal female, do not measure a mechanism (Regitz-Zagrosek et al., 2006). Sex encompasses many other qualities besides the mechanism in question and will not exhaustively describe the mechanism.

Clearly, theory often will outstrip the ability to measure, and not having access to sufficiently large cohorts and/or sufficient funds to take biological measurements may make it unfeasible to rigorously follow our proposed good practice guidelines (Giannattasio et al., 2007; Regitz-Zagrosek et al., 2006). In such cases, authors must be conservative in interpreting their findings and should avoid overbroad generalization of their findings. Studies in which sex has been used as a proxy for more specific measures can only be considered suggestive, and where possible, should be avoided.

Develop, operationalize, and test alternative hypotheses that include social and biosocial pathways

Most models in health research are incomplete and many relevant factors are omitted for practical reasons of measurement and feasibility. One way to address this problem in sex/gender health research is to develop alternative hypotheses that always include social and biosocial pathways. It is important to consider prior studies to best conceptualize and model these alternative hypotheses. For example, if prior research indicates that the size or direction of male-female difference varies over time, place, or sample characteristics (e.g., by ethnicity, age, occupation, height, menopause, or country of residence) this provides strong evidence of social and/or biosocial effects and offers insights into key variables for alternative hypotheses (Macintyre, Hunt, & Sweeting, 1996). Modeling with relevant social/behavior measures may show that the presumed biological underpinnings of the observed male-female differences in health are spurious.

Indeed, based on a plethora of evidence social causes should be assumed and a preponderance of evidence required before arguing for a primarily biological, non-social mechanism. Consider the relationship between fat intake, adiposity, and CVD (Mente, de Koning, Shannon, & Anand, 2009; Regitz-Zagrosek et al., 2006). At least two concurrent secular trends are of importance: 1) relatively high levels of unhealthy dietary fats; and 2) higher male CVD rates compared to pre-menopausal women (although the gap is narrowing) (Towfighi, Zheng, & Ovbiagele, 2009). Possible biological pathways used to explain the male-female difference focus on how men and women process fat and cholesterol in ways that differ-
associated with dietary fat and cholesterol, and provided some evidence for differential processing for males and females. But here, too, alternative hypotheses to explain the observed male-female difference using social pathways are possible. Normative conceptions of men as strong and hearty, compared to women as slender and delicate, shape the types and quantities of food eaten. Specifically, in some cultural contexts, men are more likely to eat red meat and women are more likely to eat fruits and vegetables (Prattala et al., 2007). Sometimes the richest (and most unhealthy) meat is reserved for males when there are variations in quality and/or insufficient quantity for the whole family (Holm & Mohl, 2000). Gendered social processes could be an alternative plausible mechanism linking the higher male rates of CVD with increased intake of saturated fat, or an alternative hypothesis could be based on the confluence of the social and biological mechanisms: higher male intake of fattier meats and dairy products (social processes) and differential male processing is more harmful for men (biological process).

Incorporate and test biological and social/biosocial mechanisms together

Inclusion of social and biosocial mechanisms alongside the originally proposed biological mechanism will improve design because mediation and/or spurious associations are only observable when multiple mechanisms are considered. Alternative mechanisms will help establish whether the biological mechanism is: 1) spurious and can be accounted for by social/structural factors, 2) actually a biosocial mechanism which varies in type, scope, level etc. by social/environmental causes, and/or 3) a substantively meaningful size relative to other social/structural paths. For example, biological contributions from comorbid conditions like diabetes and hypertension cannot completely account for women’s disproportionate greater adverse outcomes following an acute myocardial infarction (MI) (Naqvi et al., 2005). However, women with CVD have higher rates of depression, which, in turn, is both socially mediated and a key risk factor for poor cardiovascular outcomes (Naqvi et al., 2005; Whooley et al., 2008). The observed differential recovery from acute MI and the relatively unexplored role of depression in the causal pathway illustrates the importance of incorporating social/biosocial mechanisms within studies exploring the biological contribution to male-female differences in CVD outcomes.

A priori comparison plan

Comparison groups and statistical tests of effects should be developed based on previous empirical evidence and theory prior to proceeding with the research. Carefully planning for comparisons can help avoid the temptation to make atheoretical comparisons simply because the design includes a “sex variable” and a range of interesting outcomes. Such unplanned comparisons almost always yield false positives and can be particularly damaging given the bias towards interpreting male-female differences as ‘probably true’ and ‘probably biological’ (Jordan-Young, 2010; Kaiser et al., 2009).

Include appropriate control groups for all hypotheses (biological, social, biosocial) and accurately test for differences

Comparison groups should be strictly comparable. This is particularly important for conducting statistical tests of similarity/difference — without this strict comparability the substantive meaning of tests of statistical difference is unclear. Conclusions based on ad hoc comparison of statistical tests conducted within groups of men and women are not supportable. Phrases like “more statistically significant” in one group versus another have little practical meaning since such differences could easily be the result of differences in sample sizes, bias in assembling the population, or inadequacy of the measures used. A significant finding in one group and a non-significant finding in the other group is likewise not evidence of a male-female difference. To compare groups it is necessary to test for the significance of a statistical interaction and/or to formally test whether the effect for one group is significantly different from the effect for another group (Aiken & West, 1991).

Include sufficient measures for confounders and covariates

Statistical tests of incomplete models have limited meaning. Models must include confounders and covariates, as well as test for plausible mediation and interactions to avoid spurious findings. For example, Kanaya, Grady, and Barrett-Connor (2002) conducted a meta-analysis to assess the claim that diabetes is a stronger independent risk factor for coronary heart disease mortality in men than in women. The authors found no male-female differences after adjusting for other classic CVD risk factors (Kanaya et al., 2002). In other words, a previously established risk factor (diabetes) for male-female differences in CVD mortality was found to be incorrect when appropriate controls were included in the model.

Attend to issues of statistical significance and sample size

Sample size issues plague many studies but research often proceeds with relatively little attention to how sample size may affect findings and interpretation. The effect of sample size can operate in two directions: very large samples can produce artifactual significant relationships at the p < 0.05 level (McCloskey, 1985). Very large samples, as in gene-wide array studies, GWAS, may require a much stricter level of significance (i.e. p < 10^-5 or smaller) (Caporaso et al., 2009). One common and reasonable strategy is to use Bonferroni corrections based on the baseline p-value determined by the sample size (Rice, 1989). The Bonferroni correction simply involves dividing the baseline significance level (i.e. p < 0.05) by the number of comparisons, yielding the new p-value for statistical significance.

Small sample sizes may compromise one’s ability to detect an effect. The smaller the sample and the weaker the effect size, the more difficult it will be to achieve sufficient power. Small samples may also make testing interactions difficult if not impossible and biosocial interactions are fundamental to understanding sex/gender differences. For example, a study with 80% power to detect a main effect will only have 29% power to detect an interaction effect of the same magnitude (Brookes et al., 2004).

Conduct sensitivity analyses

To assess the robustness of the findings and to understand possible variations in effects, an array of sensitivity tests, ideally identified a priori, should be conducted. The specific tests depend on the research question(s) but could include: a) testing different operationalization of variables — e.g. a continuous vs. categorical measure of BMI; b) different modeling techniques to assess violations of modeling assumptions; c) alternative treatment of outliers, including transforming skewed variables and removing extreme cases. Sensitivity tests should be fully reported when presenting results — especially, if the outcome of sensitivity tests is not consistent with the original results.

Interpretation and presentation of results

Present within group and between group differences

Exploring variation within men and within women can help illuminate the practical/real significance of a male-female difference. If the observed magnitude of the variation within men or within women is comparable to the male-female difference, then
this difference may not be any more substantial or relevant than the natural variation within men or women. As with our earlier example, the difference in average heights between men and women in the U.S. is 5 inches (5’ft, 8 inches for men and 5’ft, 3 inches for women) (McDowell et al., 2008), but the height difference between the 5th and 95th percent of height for men and women in the U.S. is 10 inches and 9 inches respectively (McDowell et al., 2008). Such intra-sex height variations can lead to erroneous conclusions about male-female differential CVD risks. Data show that the example of reduction in arterial elasticity as a biological indicator of progression toward cardiovascular morbidity and mortality discussed earlier is dependent on height, regardless of sex/gender (Duprez et al., 2009).

Explain effect size(s): substantive vs. significant difference
Statistical significance is not substantive significance (Miller, 2005). A recent meta-analysis on sex differences in the association between birth weight and total cholesterol (a CVD risk factor) found evidence for a greater effect of low birth on cholesterol among men vs. women, attributed to “different biological processes for females and males” ((Lawlor et al. 2006) p.19.) But the effect size, 0.04 mmol/l in total cholesterol per kg birth weight, was called “unlikely to have important public health effects.” (Lawlor et al. 2006, pg 23). The substantiveness, as well as statistical significance of findings, can illuminate the real-life importance (or lack thereof) of any effect.

Report post-hoc tests and adjust significance level appropriately
Interesting findings may emerge after the hypothesized analyses, and post hoc statistical testing of these findings is good practice; however, such testing increases the likelihood of false positives. Reporting that tests were post-hoc, as well as adjusting significance level for the number of post-hoc tests assessed provides a clear account of the procedures and helps assure that significant findings are not statistical artifacts.

Describe social/biosocial attenuation and the purely social/biosocial findings
The results of the biological contribution to male–female differences in health can best be understood when situated in the context of the biosocial and social mechanism results. Describing results from all mechanisms — alternative social and biosocial mechanisms as well as full models with all mechanisms — can help illuminate the relative importance of different social, biological, and biosocial effects. This presentation of the findings allows researchers and readers to compare and theorize the potential for complementary, competing, and interactive social, biosocial, and biological mechanisms.

Describe similarities and differences in biological causes
Similarity may be larger than difference, and also very interesting and useful. Similarity should therefore be considered with the same level of sophistication and attention given to explaining differences. For example, extensive research has focused on steroid hormones as one key cause of male-female difference in CVD risk (Ding, Song, Malik, & Liu, 2006; Perez-Lopez, Larrad-Mur, Kallen, Chedraui, & Taylor, 2010). However, there are also important and interesting similarities in the effects of hormones and CVD risk for men and women. The most striking example of these male-female similarities is the narrowing of CVD risk between men and women post menopause (Mendelsohn & Karas, 2005). Research has identified several hormonal similarities that drive these findings including similar increases in CVD risk for men and post-menopausal women associated with endogenous ratios of estradiol to progesterone (He et al., 2007).

Thoroughly report limitations of the study and provide enough information for replication
Full disclosure of limitations in the design, method, or analyses of the research can help readers understand the scope and significance of the findings — as well as spur new research endeavors to help account for these limitations. Further, replication is the heart of science. Depending on space constraints and journal guidelines, it may be possible to report all information necessary to reproduce the analyses in published manuscripts. Alternatively, or in addition, manuscripts could include a link to a website that houses all information needed to reproduce the findings, including data and codes.

Discussion
We have attempted to present an alternative approach to automatic investigation of “sex differences” in human health studies and we have taken issue with assuming that sex is an individual-level variable that should be routinely assessed. Indeed, we propose the IOM mandate for male–female difference research be stood on its head: research should proceed with models based on either entirely social causes or interactive biosocial causes unless there is sound and extensive evidence for a primarily biological mechanism. We further propose that assessments of sex are almost without exception made for individuals whose development, including physical, physiological, and psychological, has been profoundly affected, in ways that we at present only dimly understand, by the operations of gender. Even when so seemingly biological and non-social an entity as a “cell” is assigned a sex, the biography of the individual that the cell came from goes along for the ride, but this is rarely acknowledged, let alone examined, in scientific analyses.

One potentially paradoxical result of our proposal to avoid the broad use of “sex” as a variable is that it could facilitate a more precise investigation of which aspects of biology contribute to male–female differences in health, and how they do so. Messing and Mager Stellman (2006), among others, have urged epidemiologists to avoid the use of the term sex as a broad proxy for biology as if the term conveyed a mechanism for generating male–female differences (Messing & Mager Stellman, 2006).

Here we used CVD to illustrate the good practice guidelines, but these guidelines are applicable to other outcomes, as supported by meta-analyses of male-female genetic differences. Patsopoulos et al. (2007) found that the vast majority of highly prominent claims of sex-related differences in GWAS were insufficiently documented or spurious. Patsopoulos and colleagues also noted that only a small fraction of supposedly a priori considerations for examining male–female differences in genetic associations did, in fact, draw upon “any kind of corroboration history” (Patsopoulos, Tatsioni & Ionnisidis 2007, 888–889). While Patsopoulos and colleagues do not make this point, we would emphasize that the habit of viewing male and female as the fundamental division between people makes it all too easy to hold a priori expectations of male–female difference even when there is no existing empirical evidence on which to base this expectation.

A final note on the long history and politics of using male–female differences in biology and health to support a wide range of political and social restrictions on women is in order (Barker-Benfield, 1976; Ehrenreich & English, 1978; Hubbard, 1990). In the 1970s, the women’s health movement noted a pervasive lack of comprehensive and unbiased knowledge about women’s bodies, and decried the tendency for medicine to proceed as if the “normal” or “ideal” patient is a white, middle-class, 160 pound man (Boston Women’s Health Book Collective, 1973; Epstein, 2007; Tuana, 2006). We recognize the IOM Report as part of a contemporary
response to the ensuing call for greater attention to women’s bodies in medical research and practice, and believe that a dedication to rectifying longstanding male biases in medicine motivates many who currently pursue sex-difference analyses. We applaud that goal, and indeed share it. But for all the reasons we have documented in this essay, we are convinced that the present approach, especially the overarching commitment to identifying and prioritizing difference over similarity, the prioritization of biology, the mandate to address sex as a grouping variable in federal funded research and clinical trials, regardless of biological evidence, and the mistaken belief that biology can be operationally separated from the social environment will not lead to the desired aim. An ever-growing catalogue of differences is not likely to achieve significant health gains for either men or women, and will only continue to distract us from the kind of focused research into specific mechanisms that we need, by offering the rather empty answer of “sex” to all questions we pose about the cause of male–female disparities.

Acknowledgments

We thank the Gender and Health Working Group at Columbia University for their insightful feedback on an earlier version of this paper, and Yoonsie Kim for her assistance with references. We also thank the Institute for Health, Health Care Policy, and Aging Research at Rutgers; the Robert Wood Johnson Health & Society Scholars Program at Columbia University; and Foundation For Worker Veteran & Environmental Health for financial support.

References


Epidemiological studies of categorical mental disorders consistently report that gender differences exist in many disorder prevalence rates and that disorders are often comorbid. Can a dimensional multivariate liability model be developed to clarify how gender impacts diverse, comorbid mental disorders? We pursued this possibility in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC; N = 43,093). Gender differences in prevalence were systematic such that women showed higher rates of mood and anxiety disorders, and men showed higher rates of antisocial personality and substance use disorders. We next investigated patterns of disorder comorbidity and found that a dimensional internalizing-externalizing liability model fit the data well, where internalizing is characterized by mood and anxiety disorders, and externalizing is characterized by antisocial personality and substance use disorders. This model was gender invariant, indicating that observed gender differences in prevalence rates originate from women and men’s different average standings on latent internalizing and externalizing liability dimensions. As hypothesized, women showed a higher mean level of internalizing, while men showed a higher mean level of externalizing. We discuss implications of these findings for understanding gender differences in psychopathology and for classification and intervention.

Keywords: comorbidity, gender differences, internalizing–externalizing, prevalence rates
largest epidemiological study of psychopathology yet undertaken (Dawson, Goldstein, Moss, Li, & Grant, 2010; Keyes, Grant, & Hasin, 2008; Grant et al., 2004; Grant & Weissman, 2007; Trull, Jahng, Tomko, Wood, & Sher, 2010; Vesga-López et al., 2008).

The origins of these gender differences in prevalence rates are not well understood although various theories have been posited to explain how they arise. These explanations include response bias, differential service utilization rates, and various biological, social, and demographic influences (see Klose & Jacob, 2004; Piccinni & Wilkinson, 2000). Psychological explanations, such as increased rumination in women partially accounting for higher rates of unipolar depression, have also been posited (Nolen-Hoeksema, 1987; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008).

These theories of gender differences focus primarily on specific disorders and rarely take comorbidity into account. A compelling account of mental disorder comorbidity focuses on unifying latent dimensional liabilities to experience multiple internalizing (mood and anxiety) or externalizing (antisocial and substance use) disorders (Eaton, South, & Krueger, 2010; Krueger, 1999; Slade & Watson, 2006; Vollebergh et al., 2001). Indeed, this internalizing–externalizing liability model is likely to frame key parts of the metastructure, or overall organization, of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–5; e.g., Andrews et al., 2009; Regier, Narrow, Kuhl, & Kupfer, 2011). The internalizing dimension can be bifurcated into distress and fear subfactors; distress relates to disorders such as major depression, dysthymia, and generalized anxiety, and fear relates to disorders such as panic disorder, social phobia, and specific phobia. The externalizing dimension is associated with disorders such as antisocial personality disorder and alcohol, nicotine, and drug dependence. Further, these factors relate to normal personality: Internalizing correlates with neuroticism/negative affectivity (Griffith et al., 2010), and externalizing correlates with disinhibition (Krueger et al., 2002).

When gender differences in prevalence rates and the internalizing–externalizing liability structure of psychopathology are considered simultaneously, the possibility of a unifying model of gender and comorbidity emerges. Specifically, women show significantly higher prevalence rates of internalizing disorders, while men show significantly higher rates of externalizing disorders (Grant & Weissman, 2007; Kessler et al., 1993, 1994). This observation suggests that gender differences in categorical prevalence rates might be due to gender differences in latent internalizing and externalizing liability dimensions. The utility of a dimensional liability model for public health, epidemiology, psychopathology, and intervention research would be notably enhanced if it could encompass the role of gender in mental disorder prevalence.

A few studies have evaluated the structure of psychopathology separately in women and men (e.g., Krueger, 1999; Kendler, Prescott, Myers, & Neale, 2003). However, we are aware of only two studies that have formally tested whether the latent structure of common mental disorders is gender invariant (Hicks et al., 2007; Kramer et al., 2008). While these studies were generally supportive of a gender invariant model, their findings’ generalizability was limited by nonrepresentative samples. Further, neither study focused on diagnoses from the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–IV; American Psychiatric Association, 1994). Addressing these limitations is critical, given the potential use of a dimensional liability model to frame key aspects of DSM–5. If the model is to be applied to DSM–5 mental disorders—and to both women and men—factorial invariance (a lack of bias) across gender should be demonstrated in a maximally representative sample.

The current study examined a nationally representative sample of 43,093 individuals assessed in the first wave of the NESARC. Our initial goal was to examine and present any gender differences in prevalence rates of common mental disorders in this sample. We were interested not only in the significance of these gender differences but also in their directionality: We hypothesized that women would show significantly higher rates of internalizing disorders than men, and men would show significantly higher rates of externalizing disorders. Second, we sought to determine the latent comorbidity structure of these disorders in women and men, separately. Third, we aimed to formally test whether or not the emergent structures in women and men could be considered gender invariant (i.e., women and men showing equivalent structures of psychopathology). Finally, if invariance were found, we hypothesized that women would have a higher mean standing on the internalizing liability dimension than men, and men would have a higher mean standing on the externalizing liability dimension. The presence of a gender invariant structure of common mental disorders would indicate that these gender differences in latent internalizing and externalizing liabilities account for the observed gender differences in prevalence rates. That is, gender differences in the prevalence of different manifest categorical disorders would be a function of mean-level gender differences in underlying liability dimensions. As such, these underlying dimensions, as opposed to their manifestations as specific observed categories of psychopathology, would be highlighted as important organizing constructs for official nosologies and for research on the role of gender in psychopathology.

Method

Participants

This study utilized data from 43,093 individuals who participated in the first wave of the NESARC, conducted in 2001–2002. The NESARC study’s design has been detailed elsewhere (Grant & Dawson, 2006). The first wave of NESARC was a representative sample of the civilian, noninstitutionalized United States population, aged 18 and older. Young adults, African Americans, and Hispanics were oversampled. Women composed 57% (n = 24,575). Race/ethnicity was selected by participants, using censusedefined categories: White (56.9%), Hispanic or Latino (19.3%), African American (19.1%), Asian/Native Hawaiian/Pacific Islander (3.1%), and American Indian/Alaska Native (1.6%). Participants provided written informed consent after reviewing a complete description of the study.

Assessment

Lifetime and past-12-month DSM–IV diagnoses were made using the Alcohol Use Disorder and Associated Disabilities Interview Schedule—DSM–IV Version (AUDADIS–IV; Grant et al., 1995), a structured interview designed for experienced lay interviewers. Major depressive disorder, dysthyemic disorder, general-
ized anxiety disorder, panic disorder, social phobia, specific phobia, alcohol dependence, nicotine dependence, marijuana dependence, other drug dependence, and antisocial personality disorder AUDADIS–IV diagnoses were examined. The other drug dependence variable was created to collapse relatively uncommon forms of drug dependence (i.e., stimulants, opioids, sedatives, tranquilizers, cocaine, solvents, hallucinogens, heroin, and any other drug not assessed) into one variable whose variance would be sufficient for covariance structure modeling; the internal consistency of this variable was good (α = .77). In keeping with DSM–IV notions of personality disorder stability, antisocial personality disorder was assessed on a lifetime basis only; this lifetime diagnosis was used in both lifetime and 12-month analyses. The reliability of the AUDADIS–IV diagnoses examined have been reported elsewhere and are generally good to excellent (e.g., kappas = .42 to .84; see Hasin et al., 2005). Test–retest estimates for AUDADIS–IV disorders are similar to other structured interviews (e.g., the Diagnostic Interview Schedule [DIS], the Composite International Diagnostic Interview [CIDI]) used in large psychiatric epidemiologic surveys (reviewed in Wittchen, 1994). Further, the AUDADIS–IV has advantages over structured interviews such as the DIS, including assessment of clinically significant distress and impairment after the syndrome is fully characterized (Hasin et al., 2005).

Statistical Analyses

All analyses were conducted in Mplus version 6 (Muthén & Muthén, 2010) using the Mplus defaults of delta parameterization and WLSMV estimator. WLSMV allowed us to treat diagnostic variables as categorical and use the NSESARC’s weighting, clustering, and stratification variables. Odds ratios used men as the reference comparison group. To evaluate model fit in confirmatory factor analyses (CFAs), we considered the comparative fit index (CFI), Tucker-Lewis index (TLI), RMSEA, and the number of freely estimated parameters in the model. CFI/TLI values > .95 and RMSEA values < .06 suggest good model fit (Hu & Bentler, 1999). Based on simulations, Cheung and Rensvold (2002) proposed that a CFI difference critical value of .01 be used to test whether the addition of constraints leads to notably worse model fit in factorial invariance studies. More parsimonious models use fewer freely estimated parameters. In model comparisons, therefore, we defined the optimal model by means of the best fit (CFI, TLI, and RMSEA), model parsimony (number of free parameters), and the CFI critical difference of .01. Distress and fear loadings on higher-order internalizing were constrained to equality to ensure model identification.

Results

Prevalence Rates

Table 1 presents the prevalence rates for the disorders included in the current study, separately for men and women and for lifetime and 12-month disorders. All odds ratios were significant (all ps < .001 except other drug dependence at p = .005). Across lifetime and 12-month prevalence rates, women showed higher rates for all internalizing disorders, and men showed higher rates for all externalizing disorders.

Comorbidity Structure

In CFAs (see Table 2), guided by previous studies and exploratory factor analyses (not reported here for brevity), we parameterized each diagnosis to load on one of three factors: (1) distress: major depression, dysthymia, and generalized anxiety disorder; (2) fear: panic disorder, social phobia, and specific phobia; and (3) externalizing: alcohol dependence, nicotine dependence, marijuana dependence, other drug dependence, other drug dependence, and antisocial personality disorder. Distress and fear were parameterized to load on a higher-order internalizing factor, which was allowed to correlate with the externalizing factor. This internalizing–externalizing model provided a very good fit in the total sample for both lifetime and 12-month diagnoses. Within each gender modeled separately, this internalizing–externalizing model continued to fit very well for lifetime and 12-month diagnoses.

Invariance

Because an internalizing–externalizing model fit well for both women and men, our next question was how similar these models

---

Table 1

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Lifetime disorders</th>
<th></th>
<th>Odds ratio</th>
<th></th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women (%)</td>
<td>Men (%)</td>
<td></td>
<td>Women (%)</td>
<td>Men (%)</td>
</tr>
<tr>
<td>Depression</td>
<td>22.9</td>
<td>13.1</td>
<td>1.46 (1.41–1.51)</td>
<td>10.1</td>
<td>5.5</td>
</tr>
<tr>
<td>Dysthymia</td>
<td>6.2</td>
<td>3.5</td>
<td>1.31 (1.25–1.38)</td>
<td>2.9</td>
<td>1.6</td>
</tr>
<tr>
<td>Generalized Anxiety</td>
<td>5.8</td>
<td>3.1</td>
<td>1.34 (1.27–1.42)</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Panic Disorder</td>
<td>7.2</td>
<td>3.7</td>
<td>1.39 (1.32–1.47)</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Social Phobia</td>
<td>5.8</td>
<td>4.3</td>
<td>1.16 (1.10–1.22)</td>
<td>3.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Specific Phobia</td>
<td>12.4</td>
<td>6.2</td>
<td>1.47 (1.41–1.53)</td>
<td>9.6</td>
<td>4.6</td>
</tr>
<tr>
<td>Alcohol Dependence</td>
<td>8.0</td>
<td>17.4</td>
<td>0.63 (0.60–0.65)</td>
<td>2.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Nicotine Dependence</td>
<td>15.6</td>
<td>20.0</td>
<td>0.84 (0.81–0.87)</td>
<td>11.5</td>
<td>14.1</td>
</tr>
<tr>
<td>Marijuana Dependence</td>
<td>0.9</td>
<td>1.7</td>
<td>0.77 (0.71–0.83)</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Other Drug Dependence</td>
<td>1.4</td>
<td>2.2</td>
<td>0.84 (0.79–0.90)</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Antisocial Personality</td>
<td>1.9</td>
<td>5.5</td>
<td>0.68 (0.59–0.66)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. All odds ratios (OR) significant at p < .001, except *p = .005. Men are OR comparison group. 95% confidence intervals are given in parentheses. Antisocial personality disorder was only assessed as a lifetime disorder.
were across gender in terms of model parameters—that is, whether the magnitude of parameters differed by gender or whether they showed invariance. Tests of invariance for indicators such as diagnoses require methodology appropriate for modeling categorical variables (Millsap & Yun-Tein, 2004). In this approach, factor loadings and thresholds are constrained to equality or freed, in tandem, across genders. In our first model (the “unconstrained model”), loadings and thresholds were free across genders, factor means were set to 0 in both genders, and scaling factors were fixed to 1 in both genders. In the second model (the “constrained model”), loadings and thresholds were constrained to equality across genders, factor means were set to 0 in men and were free in women, and scaling factors were fixed to 1 in men and were free in women. This model represented a gender invariant psychopathology structure.

We fit the unconstrained and constrained models in men and women simultaneously via a multiple group CFA, separately for lifetime and 12-month diagnoses (see Table 2). For lifetime diagnoses, the fits of the two models were identical, but the constrained model had fewer freely estimated parameters than the unconstrained model. The constrained model for lifetime diagnoses is depicted in Figure 1. For 12-month diagnoses, the constrained model had a better fit, with greater parsimony, than did the unconstrained model. For lifetime and 12-month diagnoses, the CFI critical difference of .01 was not exceeded, further supporting the constrained model. These findings indicated that, in addition to the general structure, factor loadings and thresholds for all diagnoses were equivalent for women and men. Thus, the structure of these common mental disorders, including the connections between individual diagnoses and the underlying factors, could be considered gender invariant.

In terms of factor means, means of these latent internalizing and externalizing factors were fixed to 0 in men and freely estimated as .445 and −.378 in women, respectively, for lifetime diagnoses and as .428 and −.308 for 12-month diagnoses. All mean gender differences were significant at \( p < .01 \). These standardized means

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th># Free</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample ((N = 43,093))</td>
<td>.992</td>
<td>.989</td>
<td>.012</td>
<td>—</td>
</tr>
<tr>
<td>Lifetime diagnoses</td>
<td>.988</td>
<td>.984</td>
<td>.010</td>
<td>—</td>
</tr>
<tr>
<td>Women ((n = 24,575))</td>
<td>.993</td>
<td>.991</td>
<td>.009</td>
<td>—</td>
</tr>
<tr>
<td>Lifetime diagnoses</td>
<td>.990</td>
<td>.987</td>
<td>.008</td>
<td>—</td>
</tr>
<tr>
<td>Men ((n = 18,518))</td>
<td>.988</td>
<td>.984</td>
<td>.008</td>
<td>—</td>
</tr>
<tr>
<td>Lifetime diagnoses</td>
<td>.982</td>
<td>.976</td>
<td>.007</td>
<td>—</td>
</tr>
<tr>
<td>Multigroup (Women and Men)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime diagnoses</td>
<td>.991</td>
<td>.989</td>
<td>.012</td>
<td>48</td>
</tr>
<tr>
<td>Unconstrained model</td>
<td>.991</td>
<td>.989</td>
<td>.012</td>
<td>38</td>
</tr>
<tr>
<td>Constrained model</td>
<td>.987</td>
<td>.983</td>
<td>.010</td>
<td>48</td>
</tr>
<tr>
<td>12-month diagnoses</td>
<td>.988</td>
<td>.986</td>
<td>.009</td>
<td>38</td>
</tr>
</tbody>
</table>

*Note.* Total sample analyses modeled women and men together. Multigroup analyses modeled women and men simultaneously as two separate groups. Unconstrained models allowed each gender to have unique model parameters; constrained (invariant) models constrained factor loadings and thresholds to equality across genders. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean squared error of approximation; # Free = number of freely estimated parameters.

Figure 1. The constrained (gender invariant) model in women and men using lifetime diagnoses. Values are standardized factor loadings (all significant \( p < .001 \)). Values before slash and bolded are for women; values after slash are for men. Values differ slightly across gender due to standardization. MDD = major depressive disorder; Dysth = dysthymic disorder; GAD = generalized anxiety disorder; Panic = panic disorder; Social = social phobia; Spec = specific phobia; ASPD = antisocial PD; Nic = nicotine dependence; Alc = alcohol dependence; Marj = marijuana dependence; Drug = other drug dependence. Arrows without numbers indicate unique variances, including error.
can be interpreted as $z$-scores (e.g., women were approximately .45 standard deviations higher on lifetime internalizing liability than men). Because complete factorial invariance had been established, these results demonstrated that the observed differences in the prevalence rates of the specific disorders modeled between women and men could be accounted for by the genders’ different average levels of latent internalizing and externalizing.

**Discussion**

The current study sought to synthesize two lines of research—patterns of gender differences in prevalence rates and a potentially gender invariant latent structure of psychopathology—via factorial invariance analyses of liability dimensions underlying DSM–IV disorder comorbidity. We found that the underlying structure of common mental disorders was gender invariant, with significant gender differences in mean liability levels. This provides compelling evidence that observed gender differences in prevalence rates of many common mental disorders originate at the level of latent internalizing and externalizing liabilities.

**Limitations**

This study is not without its limitations. First, we examined lifetime diagnoses, which can be subject to memory biases. However, our results from lifetime diagnoses were highly congruent with results from 12-month diagnoses, which require much less retrospection. Second, our diagnostic information was collected by extensively trained lay interviewers rather than clinicians. This being said, it is noteworthy that the instrument used to assess symptomatology was fully structured, which resulted in generally good diagnostic reliability levels. Finally, the current study investigated only common mental disorders and, thus, did not include other debilitating forms of psychopathology, such as schizophrenia. There are indications that some symptoms of psychotic disorders may relate to a separate liability factor (e.g., thought disorder liability) while also showing associations with internalizing liability/neuroticism (e.g., Barrantes-Vidal, Ros-Morente, & Kwapil, 2009; Markon, 2010).

**Implications**

**Classification.** DSM–IV is currently under revision, and there has been a great deal of discussion about the general organization of DSM–5 (Regier et al., 2011). Based largely on replications of the internalizing–externalizing model, an organizational metastructure for many common disorders reflecting this structure has been advocated (Andrews et al., 2009). Our findings support this proposal in two ways. First, we replicated the internalizing–externalizing structure in the NESARC, the largest epidemiologic study of psychopathology yet undertaken. Second, our results indicated that this structure is gender invariant. The current study represents the first time that gender invariance has been tested and successfully incorporated into the internalizing–externalizing liability model of comorbidity among categorical DSM–IV mental disorders in a representative sample. Taken together, these findings support an internalizing–externalizing metastructure for many disorders in DSM–5, especially because the model is applicable in both women and men.

**Gender differences research.** Our conclusion—that the observed gender differences in prevalence rates systematically reflect gender differences in broad latent liability factors—ties together distinct lines of research and theory on gender differences in prevalence rates for specific disorders. For instance, one major theory to account for gender differences in depression involves the notion that women ruminate more frequently than men, focusing repetitively on their negative emotions and problems rather than engaging in more active problem solving (Nolen-Hoeksema, 1987; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). This theory can be readily extended to anxiety (and other internalizing disorders) by noting that neuroticism, or negative affectivity, is strongly related to rumination such that individuals who are more neurotic ruminate more frequently (Lam, Smith, Checkley, Rijsdijk, & Sham, 2003). Neuroticism is also strongly related to ($r = .98$), and nearly isomorphic with, the latent internalizing dimension reflecting the multivariate comorbidity among DSM–IV mood and anxiety disorders (Griffith et al., 2010). This link between internalizing and trait neuroticism is itself accounted for largely by genetic effects (Hettema et al., 2006). Finally, previous research has indicated that women tend to report higher levels of trait neuroticism (as well as conscientiousness and agreeableness) on average than do men (e.g., Donnellan & Lucas, 2008), which mirrors our finding that women had significantly higher mean levels of internalizing than did men. It may be through neuroticism (and disinhibition-related traits in the case of externalizing and men; e.g., Krueger et al., 2002; Miller & Lynam, 2001; Slutske et al., 2002) that psychological processes impact latent propensities to experience comorbid mental disorders. Given that women tend to report higher frequencies of some stressful life events than men do prior to disorder onset (Harkness et al., 2010), the interaction between these liabilities and environmental stressors seems a particularly worthwhile focus for gender differences research.

**Intervention and prevention.** Our results support recent efforts to develop interventions that target latent disorder liabilities. For instance, both anxious and depressive symptoms often respond to the same pharmacologic interventions (Goldberg, Simms, Gater, & Krueger, 2011). Similarly, anxiety and depression both respond well to cognitive–behavioral therapy, and there have been efforts to develop psychotherapeutic interventions that address the shared internalizing liability rather than solely focusing on its manifestations (Barlow et al., 2011). Along these lines, prevention efforts that focus on gender-linked core psychological processes are likely to be effective in impacting multiple disorders. In women, these preventative measures might focus, for instance, on coping and cognitive restructuring skills to reduce the likelihood of rumination and cognitive distortions developing into clinically significant depression or anxiety. In men, prevention might focus on rewarding goal behaviors and shaping disinhibitory tendencies into outlets that are not destructive to the self or others.

**References**


GENDER DIFFERENCES IN PREVALENCE RATES


Received December 22, 2010
Revision received April 27, 2011
Accepted May 3, 2011
Sexual Orientation Differences in Functional Limitations, Disability, and Mental Health Services Use: Results From the 2013–2014 National Health Interview Survey

Susan D. Cochran  
University of California, Los Angeles  
Charlotte Björkenstam  
University of California, Los Angeles, and Stockholm University  
Vickie M. Mays  
University of California, Los Angeles

Objectives: The authors investigated sexual orientation differences in risk for mental health morbidity, functional limitations/disability, and mental health services use among adults interviewed in the nationally representative 2013–2014 National Health Interview Survey. Method: Respondents were 68,816 adults (67,152 heterosexual and 1,664 lesbian, gay, and bisexual [LGB] individuals), age 18 and older. Fully structured interviews assessed sexual orientation identity, health status, and services use. Using sex-stratified analyses while adjusting for demographic confounding, the authors compared LGB and heterosexual individuals for evidence of mental health-related impairments and use of mental health services. Results: LGB adults, as compared to heterosexual adults, demonstrated higher prevalence of mental health morbidity and functional limitations. However, this varied by gender with LGB women evidencing elevated risk for both mental health and substance abuse (MHSA) and non-MHSA limitations. Among men, sexual orientation differences clustered among MHSA-related limitations. Overall, LGB adults were more likely than heterosexual adults to use services, with the source of functional limitations moderating these effects among men. Conclusion: MHSA-related morbidity is a significant concern among LGB individuals and is associated with higher levels of functional limitations/disability. The findings highlight that LGB persons use MHSA-related treatment at higher rates than heterosexuals do, and, among men, are more likely to do so absent MHSA or non-MHSA-related functional limitations. This presents a unique set of concerns within the integrated care setting, including the need to deliver culturally competent care sensitive to the context of probable sex differences among LGB individuals.

What is the public health significance of this article?  
This study investigates sexual orientation differences in functional limitations, especially mental health-related, and services utilization. Lesbian, gay, and bisexual (LGB) individuals experience higher rates of mental health-related functional limitations as compared to similar heterosexual adults, with LGB women experiencing higher risk than heterosexual women for nonmental health-related limitations as well. LGB individuals are also more likely than heterosexual individuals to use mental health services and psychoactive medications regardless of their functional limitation status. Findings underscore both the possible impact of chronic stress on the lives of LGB individuals and the diversity of mental health needs within the LGB population.

Keywords: functional limitations, disability, gay, lesbian, and bisexual, mental health

Although recent changes in public opinion suggest that social stigmatization of homosexuality may be waning (Schwadel & Garneau, 2014), day-to-day discrimination remains a chronic stressor in the lives of many lesbian, gay, and bisexual (LGB) individuals (Feinstein, Goldfried, & Davila, 2012). Reflecting this, comparative studies of LGB and heterosexual individuals repeatedly

This research was supported by grants from the National Center for Minority Health and Health Disparities (MD0006923), the Wenner-Gren Foundations, and the Swedish Research Council (Vetenskapsrådet) via the Stockholm University SIMSAM Node for Demographic, Research (340-2013-5164).  
Correspondence concerning this article should be addressed to Susan D. Cochran, Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, 650 Charles E. Young Drive South, Los Angeles, CA 90095-1772. E-mail: cochran@ucla.edu
document elevated prevalence of current psychological distress, suicide attempts, major depression, generalized anxiety disorder, and other mental health/substance abuse disorders among LGB women and men (Bostwick, Boyd, Hughes, & McCabe, 2010; Cochran & Mays, 2013; Cohen, Blasey, Barr Taylor, Weiss, & Newman, 2016; Gevonden et al., 2014; Mays & Cochran, 2001; Pachankis, Cochran, & Mays, 2015). Whether these differences also translate into higher rates of functional limitations and enduring disabilities is not entirely clear (Austin, Herrick, & Proescholdbell, 2016; Björkenstam et al., 2016; Boehmer, Miao, Linkletter, & Clark, 2014; Cochran & Mays, 2007; Fredriksen-Goldsen, Kim, Barkan, Muraco, & Hoy-Ellis, 2013; Gonzales & Henning-Smith, 2015; Siordia, 2015). By functional limitations, we refer to restricted abilities to perform basic tasks (such as the ability to walk, run, or remember with or without use of special equipment) whether or not the individual seeks to use those abilities (Verbrugge & Jette, 1994). Functional limitations have the effect of reducing an individual’s capability to participate fully in daily life. In contrast, disability refers to limitations due to health-related causes in actual performance of complex behaviors and social roles, such as engaging in personal self-care (activities of daily living [ADL]), accomplishing household management tasks (instrumental activities of daily living), or, if of working age, working for gainful employment. Disability is seen as the gap between the demands of the environment and the capabilities of the individual (Verbrugge & Jette, 1994).

The current study capitalizes on the recent inclusion of sexual orientation measurement in the National Health Interview Survey (NHIS) to investigate patterns of functional limitations and disability in this potentially vulnerable population. Our work is guided by two complementary psychosocial models. The Minority Stress Model (MSM; Meyer, 2003) conceptualizes that socially marginalized identities, such as identifying as LGB, increase the risk of exposure to minority stressors, including sexual prejudice (Herek & McLemore, 2013) and discrimination. These external threats are then seen as fostering chronic hypervigilance for anticipated discrimination, internalization of negative attitudes about one’s self or one’s sexual orientation, and resulting impairments in mental well-being. To cope with minority stress, LGB individuals recruit a variety of both individual and social resources, including developing a positive gay identity, acquiring resilience skills, and creating supportive communities (Meyer, 2013). Further, we would argue that seeking counseling or psychological services to manage antigay stigma or coming out concerns is consistent with a minority stress coping response (Cochran et al., 2014; Pachankis, Hatzenbuehler, Rendina, Saffren, & Parsons, 2015). Over the years, the MSM has received substantial empirical support, particularly when predicting sexual orientation-related disparities in mental health morbidity (Cochran, 2001; Cochran & Mays, 2009; Eldahan et al., 2016; Feinstein et al., 2012; Hatzenbuehler, 2009; Leavant & Simoni, 2011; McCabe, Bostwick, Hughes, West, & Boyd, 2010; Meyer, 2003; Meyer, Schwartz, & Frost, 2008; Pachankis, Rendina, et al., 2015; Ueno, 2010). However, the MSM is less robust at explaining frequently observed patterns of gender (Cochran, Bandiera, & Mays, 2013; Cochran & Mays, 2013) or racial/ethnic (Bostwick, Boyd, Hughes, West, & McCabe, 2014; Bostwick, Meyer, et al., 2014; Calabrese, Meyer, Overstreet, Haile, & Hansen, 2015) differences within the LGB population. For example, although the MSM predicts that the double marginalization of sexual and racial/ethnic minority status should result in greater mental health disparities among LGB persons of color, current evidence is not fully supportive of this prediction (Bostwick, Meyer, et al., 2014; Calabrese et al., 2015; Cochran, Mays, Alegria, Ortega, & Takeuchi, 2007).

Thus our approach also incorporates notions drawn from the cumulative advantage/disadvantage hypothesis (CAD; Dannefer, 2003; DiPrete & Eirich, 2006). CAD refers to a systematic process whereby the effects of social advantages and disadvantages occurring earlier in life have an increasingly larger impact on individuals over the life course. These advantages arise from differences in social roles, statuses, and available resources, accumulating over time to create diversities of outcomes among individuals, even between those who might share currently similar social profiles. The integration of CAD processes brings an intersectionality approach (Dannefer, 2003) to conceptualize how differences in the life course might create diversities of outcomes within the LGB population. For example, LGB persons are less likely than heterosexual individuals to live with a relationship partner or to raise children (Carpenter & Gates, 2008). In later adulthood, single LGB individuals would then differ from single heterosexual adults in the structures of their social support networks. CAD also predicts that the impact of minority stress might vary between LGB persons of different gender, race/ethnicity, or individual backgrounds. In this regard, early mental health advantages of resiliency seen among some racial/ethnic minorities (Breslau, Kendler, Su, Gaxiola-Aguilar, & Kessler, 2005) might be protective for LGB persons of color, reducing some of the harmful impact of minority stress due to multiple disadvantaged statuses (Cochran et al., 2007).

To date, comparisons of the burden of functional limitations and disabilities between LGB and heterosexual individuals suggest that disability prevalence may be greater among LGB persons. Studies have found that LGB individuals, especially women, are more likely than heterosexual women and men to report a health disability (Austin et al., 2016; Boehmer et al., 2014; Cochran & Mays, 2007; Fredriksen-Goldsen et al., 2013; Frisch & Simonsen, 2013; Hsieh & Ruther, 2016). However, these studies did not assess the nature or extent of that disability. Additional research has compared individuals in same-sex (SS) couples to individuals in different-sex (DS) couples finding that rates of disability and receiving disability pensions are somewhat higher among persons in SS couples (Björkenstam et al., 2016; Siordia, 2015). Across existing studies, however, the extent to which mental health vulnerabilities, possibly linked to minority stress (Meyer, 2003) or CAD processes (DiPrete & Eirich, 2006), contribute to these differences is generally unknown. Gonzales and Henning-Smith (2015) did investigate mental health limitations and deficits in ADLs among coupled individuals, age 50 and older; they found that both SS coupled women and men were more psychologically distressed than DS coupled women and men, but only SS coupled women experienced excess burden for ADL deficits. Finally, for sexual minority men, HIV infection may be an important source of either disability or disabled status given the prevalence of HIV infection among gay and bisexual GB men, currently estimated at nearly one in five men (Xu, Sternberg, & Markowitz, 2010).

In the current study, we tackle three key questions. First, do LGB persons experience elevated rates of functional impairment and/or disability in everyday living? Although earlier work suggests that this may be so (Austin et al., 2016; Björkenstam et al.,
of care, nor attempted to characterize patterns of use differences that exist in the general population. For example, the effects of social marginalization associated with sexual orientation might have stronger effects on complex disabilities, such as the ability to work, or those related to social contact, but little impact on basic abilities such as ambulation.

Second, if differences do exist between sexual minority and heterosexual adults, are these differences driven primarily by mental health morbidity, reflecting the well-documented mental health disparities linked to sexual orientation status? The NHIS offers a unique opportunity to investigate this question as it includes measures of both self-reported mental and physical current health status as well as whether mental health-related symptoms and/or disorders are contributory to evident functional limitations. Because previous research has shown that mental health distress and disorders are both risk factors for and comorbid consequences of functional limitations and disability (Caputo & Simon, 2013; Shrir & Litwin, 2014), the cross-sectional design of the NHIS precludes testing causal hypotheses. However, it is possible to investigate whether sexual orientation-linked differences in functional limitations and disability are primarily clustered in the realm of mental health-linked impairments. This would be consistent with minority stress conceptualizations that emphasize the chronic nature of social marginalization in eroding one’s sense of well-being (Meyer, 2003; Meyer et al., 2008). However, there may be additional factors that influence sexual orientation-related disparities in functional limitations. Rates of tobacco use, for example, are generally higher within the LGB population when compared to heterosexual populations (Cochran et al., 2013; Corliss et al., 2013) and tobacco use is strongly linked to higher rates of functional limitations and disability (Lim et al., 2012). Thus it is possible that the nature of functional limitations and disability among LGB adults may reflect both the effects predicted by the MSM and behavioral risk differences among individuals of different sexual orientations.

Third, are there sexual orientation-related differences in patterns of mental health services utilization among those with or without functional limitations, particularly limitations attributable to mental health-related causes? Although the extant research literature indicates that LGB persons are more likely to report using mental health services (Cochran, Sullivan, & Mays, 2003), characterization of who uses services, why, or the types of services used by this population is less well understood (Flentje, Livingston, Roley, & Sorensen, 2015; Grella, Cochran, Greenwell, & Mays, 2011; Grella, Greenwell, Mays, & Cochran, 2009; McCabe, West, Hughes, & Boyd, 2013; Meyer, 2015). Some types of mental health services, especially counseling or psychotherapy, may be used by sexual minority individuals to aid in developing resiliency skills or to manage coming out issues (Pachankis, Hatzenbuehler, et al., 2015) leading to greater use of counseling or psychotherapy. However other interventions, such as psychopharmacological treatments, might not evidence sexual orientation-related use differences. To date, research on mental health services use among LGB individuals has not generally differentiated between the two types of care, nor attempted to characterize patterns of use differences that might be associated with sexual orientation. Whether the greater propensity to seek care is merely a consequence of sexual orientation-related differences in mental health morbidity or reflects distinct preferences for types of services is unknown.

Our overall goal, then, is to characterize the mental health needs of LGB persons in the United States, particularly among those with functional limitations and disability. This may assist clinicians in developing interventions within the integrated care setting to most effectively address the needs of LGB women and men who seek mental health services (Lyons, Bieschke, Dendy, Worthington, & Georgemiller, 2010).

Method

Participants and Procedure

We used publicly available data from the 2013 and 2014 NHIS (National Center for Health Statistics, 2014, 2015). This federal dataset is collected via a complex, multistage sampling design from nearly 40,000 households selected annually. In each survey year, following approved internal review board protocols, the NHIS obtains information on approximately 100,000 persons living within these households; the annual sample is representative of the civilian noninstitutionalized population in the United States. The interview process itself uses extensively trained interviewers who initially enumerate all persons and families within selected households by interviewing a household referent adult. For each family identified, a family referent adult (69% of whom are also the household referent adult) is then selected to complete a structured interview that assesses the health status of each family member including measures of functional disability. Next, one adult from each family is randomly selected to complete a more extensive health interview; all responses are self-reported. In the two, independent survey years (2013 and 2014), household response rates for initial screening averaged 75%; conditional response rates for the sampled adult interview averaged 81%. Beginning in the 2013 survey year, the NHIS added a single question to the sampled adult interview assessing sexual orientation identity. Across the 2013–2014 surveys, 71,254 adults, aged 18 years and older, completed the sampled adult interview. Of these, 69,718 were administered the sexual orientation question with 68,816 providing usable responses. These individuals, referred to henceforth as respondents, comprise the study sample. Because of the public nature of the NHIS dataset, the current study was exempt from further IRB review requirements.

Measures

Sexual orientation. Respondents were asked “Which of the following best represents how you think of yourself?” Four answers were offered: (a) lesbian or gay (for women) and gay (for men); (b) straight (i.e., not lesbian or gay); (c) bisexual; or (d) something else. Persons selecting something else (n = 144) provided no further information and were dropped from the sample. We also excluded those who refused to respond (n = 448) or indicated they did not know how to answer the question (n = 310). For the remaining respondents, we categorized sexual orientation into two groups: those identifying as heterosexual (n = 67,152) or
as sexual minority (n = 1,664), including 1,149 who identified as lesbian or gay and 515 as bisexual.

**Functional limitations and disability.** The NHIS extensively measured functional limitations (Brandt, Ho, Chan, & Rasch, 2014). This was done both in the family interview, conducted with the family referent adult (71% of whom were also the sampled adult respondent) and in the sampled adult interview. In the family interview, the referent was asked if the sampled adult (respondent) currently experienced limitations or difficulties in functioning in any of several domains due to “a physical, mental or emotional problem.” The functional domains assessed were walking; remembering things; ADL, such as bathing, dressing, or getting around; and instrumental activities of daily living such as doing household chores or shopping. Respondents with one or more areas of reported functional difficulty were classified by NHIS as having a general functional activity limitation. A subsequent question asked what medical conditions or health problems caused these limitations. NHIS coded open-ended responses into one of 37 health conditions including three mental health/substance abuse-related (MHSA) causes: depression, anxiety, or emotional problems (corresponding to ICD-9 CM Codes: 300.0–302.9, 306–313.9); alcohol, drugs, or substance use (ICD-9 CM codes: 291–292.9, 303–305.9); and any other mental disorder (ICD-9 CM codes: 290–290.9, 293–299.9, 314.00, 314.01). We further classified general functional activity limitations into one of two types: those in which MHSA-related conditions were contributory and those in which MHSA-related conditions were absent. In addition, the family referent adult rated the respondent’s current general health status. This we collapsed into one of two categories (excellent, very good, good vs. fair, poor). For respondents aged 18 to 64 years, the family referent also indicated whether the respondent received disability-related income from Social Security, the railroad retirement system, or other sources. This we used to classify respondents as receiving disability income or not.

Functional limitations were also assessed during the sampled adult interview by asking respondents whether they experienced difficulties, due to a health problem, in performing 12 activities. These were drawn from two domains: (a) mobility deficits (e.g., reaching, stooping, standing, climbing, walking, sitting, grasping, carrying, and pushing) and (b) participating in common life activities (e.g., going out to shop, seeing a movie, or attending a sporting event; going out to visit friends or attend social gatherings; and engaging in relaxing pursuits such as reading, listening to music, or watching TV at home). From this, we created three measures: (a) any mobility deficits reported, (b) any life participation limitation reported, and (c) a summary measure capturing reports of either or both mobility and/or life participation limitations. Respondents also indicated up to three medical conditions or health problems that caused these limitations. NHIS coded these into 37 categories identical to those listed above. Limitations in the survey question structure prevented us from classifying the reasons for mobility and life participation limitations separately. However, we were able to classify the summary measure of mobility/life participation limitations into those in which MHSA-related conditions were contributory versus those in which they were not. Finally, we created a summary variable to capture whether any functional limitations, both MHSA-related and non-MHSA-related, were reported either in the family level interview or the sampled adult interview.

**Mental health indicators.** The sampled adult interview included the K-6 Distress Scale, a mental health screening tool for nonclinical populations (Kessler et al., 2002). The K-6 measures nonspecific distress occurring in the 30 days prior to interview by assessing the frequency with which individuals have feelings of sadness, nervousness, restlessness, hopelessness, worthlessness, and that everything is an effort. Scores range from 0 to 24; a standard cutoff score for high psychological distress (scores of 13 or greater) identifies individuals at high risk of meeting diagnostic criteria for serious mental illness. One additional question in the K-6 assesses the extent to which these feelings “interfere with life or activities” (not at all, a little, some, a lot). We recoded this item into two categories (no interference vs. any interference). Respondents were also asked if they had seen a mental health professional (a psychiatrist, psychologist, psychiatric nurse, or clinical social worker) in the 12 months prior to interview which we coded as yes or no.

Finally, NHIS selected a random half of sampled adult respondents to assess the frequency (a few times a year, monthly, weekly, or daily) with which they experienced feelings of anxiety and/or depression and the intensity of each of these feelings the last time they occurred (a little, a lot, or somewhere in between a little and a lot). To approximate a more clinically relevant measure, we created two separate composite variables for both anxiety and depression that classified individuals into one of two groups: those who reported experiencing the mood state at least monthly with high intensity (a lot) and those who did not. These measures were moderately correlated with high psychological distress as indexed by the K-6 (for anxiety, r = .41, p < .001; for depression, r = .45, p < .001). Two additional variables asked respondents (yes, no) if they were taking medication to manage feelings of anxiety and depression. From these responses, we created two variables: takes antidepressant drug (yes, no), takes antianxiety drug (yes, no).

Finally, we also created two summary variables. The first identified individuals reporting any anxiety and/or depression with monthly frequency and high intensity. The second captured whether the respondent was taking medication to manage their symptoms.

**Demographic characteristics and health insurance status.** The interview assessed respondents’ gender, age, race/ethnicity, country of birth, relationship status, educational attainment, family income, region of residence, and current health insurance status. We coded age into one of three categories (18–34 years, 35–54 years, and 55 years or older), race/ethnicity into two (non-Hispanic White vs. all other racial/ethnic backgrounds), country of birth into two (U.S. born vs. foreign born), educational attainment into two (high school degree or less vs. some college or more), and relationship status into two (married or living with partner vs. single,widowed, separated or divorced). Family income, weighted by family size, was categorized by NHIS as either below 200% of the federal poverty level or equal to or greater than 200% federal poverty level. The NHIS data sets also provide information on residential location using standard U.S. Census region classifications (Northeast, Midwest, South, and West). Finally, the NHIS summarized their extensive assessment of health insurance coverage by creating a single variable to identify individuals lacking health insurance coverage at the time of interview.
Analysis

Data were analyzed using Stata (Stata Corporation, 2013) employing both survey design information and weights as advised by the National Center for Health Statistics. Survey design information appropriately inflates standard errors to take into account the greater uncertainty of estimates derived from clustered sampling designs. The sampling weight adjusts for selection probability (e.g., single adults are more likely to be sampled than adults living in multiple adult households), nonresponse, and poststratified adjustments of the obtained sample to more closely match the age, gender, and race/ethnicity composition of the U.S. population. Because sexual orientation effects often differ by gender (Cochran & Mays, 2007, 2009) and women and men often experience different patterns of functional limitations (Liang et al., 2008), we adopted a sex-stratified analytic approach. Initial analyses investigated possible demographic differences between heterosexual and LGB women and men separately using logistic regression methods that regressed sexual orientation on age, race, country of birth, relationship status, educational attainment, family income, geographic region, and survey cycle simultaneously. Next, we used logistic and multinomial analyses to investigate sex-stratified sexual orientation differences in insurance status, general health, distress, and functional limitations/disabilities. Here, we adjusted for possible confounding due to the demographic variables listed above as well as survey cycle with one exception: analyses of disability-related income did not adjust for family income. Next, we used similar methods to examine sexual orientation-related differences in mental health treatment utilization, with additional control of confounding due to health insurance status. Finally, to investigate sexual orientation differences in experiences with frequent anxiety and depression, we restricted our sample to those respondents who completed the module wherein these questions were asked. In these analyses, three approaches were used. For mood frequencies, we regressed reports of mood experiences on sexual orientation while adjusting for demographic confounding and survey cycle. For analyses investigating health care and medication use, we additionally adjusted for health insurance status. In addition, to assess effect modification of limitation status, we also evaluated the possible interaction between functional limitation status (evidence of any reported limitation in either the family or sample adult interview) and sexual orientation in use of mental health services.

Because LGB respondents were less likely to be partnered than heterosexual respondents, they were also somewhat more likely to be selected as both the family referent and the sampled adult (65.7% of LGB respondents vs. 60.8% of heterosexual respondents), Wald $F(1) = 7.49, p < .01$. Hence, we repeated our analyses restricting the dataset to respondents for whom the family referent was the sampled adult. Sexual orientation-linked differences that rose to the level of statistical significance were identical to those obtained in the full sample with one exception (gay and bisexual men were more likely than heterosexual men to rate their health as fair or poor in the restricted sample, results available from the authors). For brevity, we only report findings from the total sample. Below, we provide weighted crude prevalences and their standard errors (SEs); means (Ms), results of Wald $F$ tests adjusted (adj.) for confounders where appropriate; and adjusted relative risk ratios (RRs) and their confidence intervals (CIs). Because functional limitations are strongly related to age, as was sexual orientation, we also report predicted marginal prevalences for disability-related outcomes that adjust for demographic differences between LGB and heterosexual respondents. Significance tests were based on the criterion of $p < .05$; all confidence intervals were estimated with 95% certainty.

Results

Demographic Correlates of Sexual Orientation

Overall, 2.3% (CI[2.2–2.5%]) of respondents identified as lesbian, gay, or bisexual, including 2.4% (CI[2.2–2.6%]) of women and 2.3% (CI[2.0–2.5%]) of men. Among women, LGB women differed from heterosexual women on a number of demographic characteristics, including being somewhat younger ($M_{age} = 38.4$ years for sexual minority women vs. $M_{age} = 47.7$ for heterosexual women), adj. Wald $F(1) = 41.35, p < .001$; more likely U.S. born, adj. Wald $F(1) = 6.56, p = .01$; and less likely to be married or cohabiting, adj. Wald $F(1) = 12.80, p < .001$; as well as reporting lower family income, adj. Wald $F(1) = 4.23, p < .05$; and not living in the Midwest, adj. Wald $F(1) = 2.92, p < .05$ (see Table 1). Similarly, among men, sexual minority men differed from heterosexual men in that they were somewhat younger ($M_{age} = 42.2$ years for sexual minority men vs. $M_{age} = 46.2$ for heterosexual men), adj. Wald $F(1) = 9.09, p < .01$; more likely U.S. born, adj. Wald $F(1) = 21.08, p < .001$; less likely to be married or cohabiting, adj. Wald $F(1) = 110.77, p < .001$; possessed more education, adj. Wald $F(1) = 40.82, p < .001$; and were less likely to be living in the Midwest or South, adj. Wald $F(1) = 4.55, p < .01$. However, sexual minorities, both women, adj. Wald $F(1) = 1.96, p = .16$, and men, adj. Wald $F(1) = 0.21, p = .64$, did not differ significantly from heterosexual women or men, respectively, in their levels of health insurance coverage.

Health Status and Functional Limitations

Among women, lesbians and bisexuals were more likely than heterosexuals to evidence high levels of recent psychological distress, adj. Wald $F(1) = 17.93, p < .001$, after adjusting for confounding (see Table 2). They were also more likely to report that psychological distress interfered with their lives, adj. Wald $F(1) = 24.42, p < .001$; that their general health status was somewhat worse, adj. Wald $F(1) = 7.67, p < .01$; and that they had seen a mental health provider in the prior year, adj. Wald $F(1) = 41.57, p < .001$. This higher level of psychological morbidity among LGB women was reflected in more frequent complaints of functional limitations than that seen among heterosexual women—functional limitations reported by the family referent, adj. Wald $F(1) = 16.82, p < .001$; mobility/life participation limitations reported by the respondent, adj. Wald $F(1) = 17.51, p < .001$; and any functional limitations reported: adj. Wald $F(1) = 19.56, p < .001$. The types of deficits LGB women reported were not limited to complex social engagement abilities. For example, prevalences of mobility impairments, adjusted for demographic differences, indicated that approximately 48% (CI[43–52%]) of sexual minority women experienced mobility impairments compared to 38% (CI[37–39%]) of heterosexual women (adj. RR = 1.26, CI[1.14–1.38]). Similar comparisons of problems in life participation demonstrated...
### Table 1
Characteristics of Respondents in 2013–2014 National Health Interview Survey by Gender and Sexual Orientation: Weighted Prevalences and Results of Logistic Regression Models Shown

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sexual minority</td>
<td>Heterosexual</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–34 years</td>
<td>48.5 (.24)</td>
<td>29.0 (.4)</td>
</tr>
<tr>
<td>35–54 years</td>
<td>35.4 (.22)</td>
<td>34.2 (.3)</td>
</tr>
<tr>
<td>55 years or older</td>
<td>16.0 (1.7)</td>
<td>36.8 (.4)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>68.0 (2.2)</td>
<td>66.2 (.4)</td>
</tr>
<tr>
<td>Location of residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>25.8 (2.0)</td>
<td>22.3 (.4)</td>
</tr>
<tr>
<td>Northeast</td>
<td>17.9 (2.1)</td>
<td>17.3 (.3)</td>
</tr>
<tr>
<td>Midwest</td>
<td>19.2 (1.8)</td>
<td>22.5 (.4)</td>
</tr>
<tr>
<td>South</td>
<td>37.2 (2.2)</td>
<td>37.9 (.4)</td>
</tr>
<tr>
<td>Lacks health insurance</td>
<td>18.3 (1.8)</td>
<td>13.3 (.2)</td>
</tr>
</tbody>
</table>

**Note.** Sexual minorities included 878 lesbian, gay, and bisexual women, 786 gay and bisexual men; heterosexuals included 37,185 women and 29,967 men. RR = relative risk; CI = confidence interval; FPL = Federal Poverty Limit; Ref = referent. Demographic differences evaluated by logistic regression methods regressing sexual orientation on all characteristics (gender, age, race/ethnicity, nativity, relationship status, education, family income, geographic region, and survey cycle) simultaneously. Differences in health insurance coverage evaluated by logistic regression methods regressing insurance status on sexual orientation, other demographic characteristics, and survey cycle simultaneously.

Adjusted prevalences of 16% (CI[13–21%]) among sexual minority women versus 12% (CI[12–13%]) among heterosexual women (adj. RR = 1.36, CI[1.07–1.73]). These sexual orientation-related differences were not restricted to limitations attributed to MHSA-related causes. Reflecting the elevated disability risk, among women of working age (age 18 to 64 years), LGB women were more likely than similar heterosexual women to be receiving disability pensions, adj. Wald $F(1) = 10.38, p < .01$.

### Table 2
Indicators of Distress, Functional Limitations, and Mental Health Services Use Among Women, Age 18 and older, in 2013–2014 National Health Interview Survey by Sexual Orientation: Crude and Adjusted Prevalences and Partial Results of Regression Models Shown

<table>
<thead>
<tr>
<th>Morbidity indicators</th>
<th>Crude prevalence</th>
<th>Adjusted prevalence</th>
<th></th>
<th></th>
<th></th>
<th>Adjusted RR [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sexual minority</td>
<td>Heterosexual</td>
<td>Sexual minority</td>
<td>Heterosexual</td>
<td>Adjusted RR</td>
<td></td>
</tr>
<tr>
<td>Morbidity indicators</td>
<td>% (SE)</td>
<td>% (SE)</td>
<td>% (SE)</td>
<td>% (SE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High psychological distress, past 30 days</td>
<td>7.5 (1.0)</td>
<td>3.8 (.1)</td>
<td>7.2 (1.0)</td>
<td>3.8 (.1)</td>
<td>1.88 [1.41–2.50]</td>
<td></td>
</tr>
<tr>
<td>Psychological distress interferes with life</td>
<td>33.7 (2.2)</td>
<td>21.4 (.3)</td>
<td>31.3 (2.2)</td>
<td>21.5 (.3)</td>
<td>1.46 [1.27–1.68]</td>
<td></td>
</tr>
<tr>
<td>Overall health status rated fair/poor</td>
<td>14.5 (1.7)</td>
<td>13.4 (.2)</td>
<td>18.3 (2.0)</td>
<td>13.3 (.2)</td>
<td>1.38 [1.11–1.71]</td>
<td></td>
</tr>
<tr>
<td>General functional limitations (family referent reported)</td>
<td>18.5 (1.7)</td>
<td>8.1 (.2)</td>
<td>15.6 (1.4)</td>
<td>8.1 (.2)</td>
<td>1.92 [1.59–2.32]</td>
<td></td>
</tr>
<tr>
<td>Mobility deficits/life participation limitations (respondent reported)</td>
<td>11.0 (1.5)</td>
<td>13.8 (.2)</td>
<td>15.8 (1.9)</td>
<td>13.7 (.2)</td>
<td>1.15 [1.91–1.46]</td>
<td></td>
</tr>
<tr>
<td>Receives disability income (18–64 years only)</td>
<td>9.2 (1.5)</td>
<td>6.3 (.2)</td>
<td>10.7 (1.7)</td>
<td>6.2 (.2)</td>
<td>1.71 [1.25–2.35]</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** N = 38,063, including 878 lesbian, gay, and bisexual women; 37,185 heterosexual women. RR = relative risk; CI = confidence interval; MHSA = mental health and substance abuse. Predicted marginals (adjusted prevalences) and adjusted RR were obtained by logistic or multinomial logistic regression methods evaluating sexual orientation differences while adjusting for confounding due to demographic characteristics (age, race/ethnicity, nativity, relationship status, education, family income, geographic region) and survey cycle. Mental health provider treatment differences were further adjusted for health insurance status while disability income was not adjusted for family income.
In contrast, among men, sexual orientation differences were most strongly clustered with markers of mental health morbidity. As was seen among women, gay and bisexual men were more likely than heterosexual men to evidence high levels of psychological distress, adj. Wald $F(1) = 22.36, p < .001$; to indicate that psychological distress interfered with their lives, adj. Wald $F(1) = 29.41, p < .001$; and to report having seen a mental health provider in the prior year, adj. Wald $F(1) = 63.97, p < .001$ (see Table 3). Further, sexual orientation differences were present in measures of functional limitations, including general functional limitations, adj. Wald $F(1) = 4.73, p < .01$; mobility/life participation limitations, adj. Wald $F(1) = 10.30, p < .001$; and any functional limitations, adj. Wald $F(1) = 7.43, p < .001$. However, in each case the differences between sexual minority and heterosexual men lay in greater prevalence of limitations that were attributed to MHSA-related causes. Consistent with these findings, sexual minority men were no more likely than heterosexual men to report fair or poor health, adj. Wald $F(1) = 1.21, p = .27$; mobility deficits, adj. RR = 1.14, CI[0.99–1.31]; or, among men of working age, to report receiving disability pensions, adj. Wald $F(1) = 0.42, p = .52$. However, gay and bisexual men were more likely to report limitations in life participation activities, adj. RR = 1.44, CI: 1.08–1.92.

### Prominence of Mood-Related Differences

Approximately 52% (CI[49.1–55.6%]) of LGB respondents and 50% (CI[49.3–50.4%]) of heterosexual respondents completed the brief adult disability module, Wald $F(1) = 2.12, p = .14$. Here, too, there was strong evidence that negative affect was more frequently burdensome for sexual minority persons than for heterosexual persons (see Table 4). Specifically, among women, LGB respondents were more likely than heterosexual women to report experiencing monthly high intensity anxiety, adj. Wald $F(1) = 15.40, p < .001$; depression, adj. Wald $F(1) = 13.62, p < .001$; and either or both mood states, adj. Wald $F(1) = 15.62, p < .001$. Similarly, among men, sexual minority men were more likely than heterosexual men to report high intensity, monthly anxiety, adj. Wald $F(1) = 42.25, p < .001$; depression, adj. Wald $F(1) = 16.03, p < .001$; and either or both mood states, adj. Wald $F(1) = 43.71, p < .001$. Medication use to treat these symptoms was also associated with sexual orientation. LGB women were more likely than heterosexual women to be taking medications to treat anxiety, adj. Wald $F(1) = 8.90, p < .01$; depression, adj. Wald $F(1) = 17.80, p < .01$; and/or both symptoms, adj. Wald $F(1) = 16.18, p < .001$; among men, gay and bisexual men were more likely than heterosexual men to be using medication to treat anxiety, adj. Wald $F(1) = 31.87, p < .001$; depression, adj. Wald $F(1) = 38.60, p < .001$; and/or both symptoms, adj. Wald $F(1) = 32.25, p < .001$.

Use of mental health services was also strongly and positively related to functional limitations. Although only 11% (CI[10.4–12.0%]) of those individuals without any functional limitations who completed the adult disability module reported use of mental health services, including taking medications to treat anxiety and/or depression or seeing a mental health specialist, nearly 22% (CI[20.4–23.0%]) of those with a non-MHSA-related functional limitation had utilized care, and fully 76% (CI[71.6–80.0%]) with a MHSA-related limitation had done so. This use was also associated with sexual orientation status, however tests evaluating the interaction between sexual orientation and functional limitations showed no significant interactions.

### Table 3

<table>
<thead>
<tr>
<th>Morbidity indicators</th>
<th>Crude prevalence</th>
<th>Adjusted prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sexual minority</td>
<td>Heterosexual</td>
</tr>
<tr>
<td>High psychological distress, past 30 days</td>
<td>7.5 (1.5)</td>
<td>2.8 (.1)</td>
</tr>
<tr>
<td>Psychological distress interferes with life</td>
<td>28.3 (2.2)</td>
<td>16.4 (.3)</td>
</tr>
<tr>
<td>Overall health status rated fair/poor</td>
<td>11.8 (1.4)</td>
<td>12.3 (.2)</td>
</tr>
<tr>
<td>Saw mental health provider, past year</td>
<td>20.5 (2.4)</td>
<td>5.8 (.2)</td>
</tr>
<tr>
<td>General functional limitations (family referent reported)</td>
<td>10.9 (1.5)</td>
<td>12.3 (.2)</td>
</tr>
<tr>
<td>Any—not MHSA-related</td>
<td>5.3 (1.2)</td>
<td>2.2 (.1)</td>
</tr>
<tr>
<td>Mobility deficits/life participation limitations (respondent reported)</td>
<td>25.6 (2.1)</td>
<td>27.1 (.4)</td>
</tr>
<tr>
<td>Any—not MHSA-related</td>
<td>5.5 (1.3)</td>
<td>1.6 (.1)</td>
</tr>
<tr>
<td>Any limitations reported by either source</td>
<td>26.2 (2.1)</td>
<td>28.2 (.4)</td>
</tr>
<tr>
<td>Any—not MHSA-related</td>
<td>7.5 (1.5)</td>
<td>2.9 (.1)</td>
</tr>
<tr>
<td>Receives disability income (18–64 years only)</td>
<td>8.1 (1.4)</td>
<td>6.9 (.2)</td>
</tr>
</tbody>
</table>

Note. $N = 30,753$, including 786 gay and bisexual men and 29,967 heterosexual men. RR = relative risk; CI = Confidence interval; MHSA = mental health and substance abuse. Predicted marginals (adjusted prevalences) and adjusted RR were obtained by logistic or multinomial logistic regression methods evaluating sexual orientation differences while adjusting for confounding due to demographic characteristics (age, race/ethnicity, nativity, relationship status, education, family income, geographic region) and survey cycle. Mental health provider treatment differences were further adjusted for health insurance status while disability income was not adjusted for family income.
Differences while adjusting for confounding due to demographic characteristics (age, race/ethnicity, nativity, relationship status, education, family income, geographic region) and survey cycle. Differences in medication and health services use were further adjusted for health insurance status.

Smith, 2015; Hsieh & Ruther, 2016; Siordia, 2015), LGB respondents in the NHIS evidenced higher prevalence of functional limitations and disability than did heterosexual respondents. Capturing on NHIS’s extensive measurement of limitations and social engagement that LGB individuals reported were associating with MHSA-related factors. The breadth of these findings underscore that social marginalization of LGB individuals likely has a costly, chronic and harmful impact.

At the same time, we also found evidence of important differences in the patterns of these effects when analyzed separately for men and women. Among men, for example, sexual orientation-related variations in mobility limitations were not prominent, and functional limitations and disabilities were restricted to MHSA-related causes. In contrast, among women, sexual orientation-related differences were seen in measures of both mobility and life participation and were linked to both physical health and MHSA-related causes. Although this difference in sexual orientation-related health disparities is not well predicted by the MSM (Meyer, 2003), it is consistent with CAD hypotheses (Dannefer, 2003; DiPrete & Eirich, 2006) that emphasize differential impact of the same stressors depending on earlier patterns of individual and social status advantages and disadvantages. To that end, emerging evidence suggests that sexual minority men and women experience somewhat different responses and trajectories of health challenges across their lifespans. For example, although a sizable minority of gay and bisexual men must cope with prevalent HIV infection (Xu et al., 2010), sexual minority men, on the whole, are more physically fit than heterosexual men (Cochran, Björkenstam, & Mays, 2013; Cochran & Mays, 2007; Fredriksen-Goldsen et al., 2013; Gonzales & Henning-Smith, 2015; Hsieh & Ruther, 2016; Siordia, 2015), LGB respondents in the NHIS evidenced higher prevalence of functional limitations and disability than did heterosexual respondents. Capitalizing on NHIS’s extensive measurement of limitations and disability, we were also able to extend these findings by showing that much of the health disadvantages in mobility, independence, and social engagement that LGB individuals reported were associated with MHSA-related factors. The breadth of these findings underscore that social marginalization of LGB individuals likely has a costly, chronic and harmful impact.

**Discussion**

Consistent with both predictions from the MSM (Meyer, 2003) and results from earlier studies (Bostwick et al., 2010; Cochran & Mays, 2013; Cohen et al., 2016; Gevonden et al., 2014; Pachankis, Cochran, et al., 2015), we found strong evidence that LGB individuals, when compared to similar heterosexual adults, report higher current rates of depressed mood, anxiety, and psychological distress. Also, as others have observed (Austin et al., 2016; Björkenstam et al., 2016; Boehmer et al., 2014; Cochran & Mays, 2007; Fredriksen-Goldsen et al., 2013; Gonzales & Henning-Smith, 2015; Hsieh & Ruther, 2016; Siordia, 2015), LGB individuals in the NHIS evidenced higher prevalence of functional limitations and disability than did heterosexual respondents. Capturing on NHIS’s extensive measurement of limitations and social engagement that LGB individuals reported were associated with MHSA-related factors. The breadth of these findings underscore that social marginalization of LGB individuals likely has a costly, chronic and harmful impact.

At the same time, we also found evidence of important differences in the patterns of these effects when analyzed separately for men and women. Among men, for example, sexual orientation-related variations in mobility limitations were not prominent, and functional limitations and disabilities were restricted to MHSA-related causes. In contrast, among women, sexual orientation-related differences were seen in measures of both mobility and life participation and were linked to both physical health and MHSA-related causes. Although this difference in sexual orientation-related health disparities is not well predicted by the MSM (Meyer, 2003), it is consistent with CAD hypotheses (Dannefer, 2003; DiPrete & Eirich, 2006) that emphasize differential impact of the same stressors depending on earlier patterns of individual and social status advantages and disadvantages. To that end, emerging evidence suggests that sexual minority men and women experience somewhat different responses and trajectories of health challenges across their lifespans. For example, although a sizable minority of gay and bisexual men must cope with prevalent HIV infection (Xu et al., 2010), sexual minority men, on the whole, are more physically fit than heterosexual men (Cochran, Björkenstam, & Mays, 2013; Cochran & Mays, 2007; Fredriksen-Goldsen et al., 2013; Gonzales & Henning-Smith, 2015; Hsieh & Ruther, 2016; Siordia, 2015), LGB respondents in the NHIS evidenced higher prevalence of functional limitations and disability than did heterosexual respondents. Capturing on NHIS’s extensive measurement of limitations and disability, we were also able to extend these findings by showing that much of the health disadvantages in mobility, independence, and social engagement that LGB individuals reported were associated with MHSA-related factors. The breadth of these findings underscore that social marginalization of LGB individuals likely has a costly, chronic and harmful impact.

**Table 4**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sexual minority</th>
<th>Heterosexual</th>
<th>Adjusted RR [95% CI]</th>
<th>Sexual minority</th>
<th>Heterosexual</th>
<th>Adjusted RR [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms and medication use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiences anxiety</td>
<td>13.0 (2.1)</td>
<td>5.7 (2.3)</td>
<td>1.98 [1.42–2.75]</td>
<td>12.6 (2.4)</td>
<td>3.5 (2)</td>
<td>3.57 [2.50–5.12]</td>
</tr>
<tr>
<td>Takes anxiety medication</td>
<td>17.0 (2.3)</td>
<td>11.2 (.3)</td>
<td>1.51 [1.16–1.97]</td>
<td>16.2 (2.5)</td>
<td>5.7 (.3)</td>
<td>2.62 [1.91–3.60]</td>
</tr>
<tr>
<td>Experiences depression</td>
<td>9.7 (1.8)</td>
<td>4.2 (.2)</td>
<td>2.08 [1.43–3.03]</td>
<td>6.7 (1.5)</td>
<td>2.6 (.2)</td>
<td>2.49 [1.61–3.83]</td>
</tr>
<tr>
<td>Takes depression medication</td>
<td>18.0 (2.4)</td>
<td>10.2 (.3)</td>
<td>1.81 [1.39–2.35]</td>
<td>17.1 (2.8)</td>
<td>5.2 (.3)</td>
<td>3.05 [2.20–4.23]</td>
</tr>
<tr>
<td>Experiences either mood</td>
<td>16.4 (2.2)</td>
<td>7.4 (.3)</td>
<td>1.96 [1.48–2.60]</td>
<td>14.4 (2.4)</td>
<td>4.5 (.2)</td>
<td>3.14 [2.29–4.32]</td>
</tr>
<tr>
<td>Takes medication for either</td>
<td>22.1 (2.7)</td>
<td>13.6 (.3)</td>
<td>1.63 [1.30–2.05]</td>
<td>19.5 (2.8)</td>
<td>7.0 (.3)</td>
<td>2.55 [1.89–3.45]</td>
</tr>
</tbody>
</table>

**Note.** Respondents were randomly selected to complete the Functioning and Disability module of the Sample Adult Survey. RR = relative risk; CI = Confidence interval; MHSA = mental health and substance abuse. Adjusted RR was obtained by logistic regression methods evaluating sexual orientation differences while adjusting for confounding due to demographic characteristics (age, race/ethnicity, nativity, relationship status, education, family income, geographic region) and survey cycle. Differences in medication and health services use were further adjusted for health insurance status.

* Person reports experiencing anxiety or depression of high intensity at least monthly.  * Person currently takes medication for anxiety and/or depression.
important additional predictors of sexual orientation-related disparities in health.

Importantly, our study also investigated sexual orientation-related differences in patterns of mental health services use among those with or without functional limitations. Like others, we found that LGB individuals have a greater propensity to use mental health services than heterosexual women and men do (Bostwick et al., 2010; Cochran & Mays, 2013; Pachankis, Cochran, et al., 2015). However, here, too, we also observed hints of possible sex differences in patterns of services use among LGB individuals. For example, among LGB women, as compared to heterosexual women, the greater propensity to use services is evident among those with and without functional limitations. However, for men sexual orientation differences were greatest among those men without MHSA-related functional limitations. Whether this reflects treatment seeking for assistance with stigma-related issues (Pachankis, Hatzenbuehler, et al., 2015) is indeterminable but worthy of future research.

Three limitations warrant comment in contextualizing these results. First, because of sample size limitations we were unable to investigate heterogeneity of effects within sexual minority subgroups (e.g., gay vs. bisexual). Elsewhere (Bostwick et al., 2010; Cochran & Mays, 2013; Fredriksen-Goldsen et al., 2013; Pachankis, Cochran, et al., 2015), research strongly suggests that these differences in effects likely exist. Only with additional years of data collection in the annually mounted NHIS will the dataset be capable of supporting such work. Second, the NHIS does not assess the content of services provided by mental health providers or the provision of mental health services by non-specialists, such as primary care physicians. This has two implications. First, although we have inferred that sexual minorities may be using mental health services to address issues of coping with discrimination or to develop resiliency skills, we have no means of evaluating the validity of that inference. Second, if there are sexual orientation-linked differences in seeking assistance from different types of providers our results will be biased in indeterminable ways. For example, if LGB individuals are more likely than heterosexual persons to receive treatment from primary care physicians, we will have underestimated sexual orientation differences. A third study limitation is that the NHIS is a health surveillance survey for the general population and does not include measurement of constructs, such as perceived discrimination or minority stress. Although we have inferred that the sexual orientation-related disparities observed here arise from the pernicious effects of social adversity and the ways in which it shapes, through cumulative advantages and disadvantages, the lives of LGB persons, a direct test of that hypothesis is not possible with the current dataset.

Nevertheless, our findings strongly support the perspective that social adversity serves as a chronic stressor for a sizable minority of LGB individuals (Cochran, 2001; Cochran & Mays, 2009; Feinstein et al., 2012; Hatzenbuehler, 2009; Lehavot & Simoni, 2011; Meyer, 2003; Meyer et al., 2008). Our findings also bring some clarity to these chronic health disparities suggesting that gay and bisexual men seem to experience a focal MHSA-related burden while lesbian and bisexual women evidence both physical and mental health-related limitations. Whether this reflects differential effects of minority stress or arises from structural differences in the lives of LGB women and men borne of their life course experiences is currently unknown.

Given the sexual orientation-linked disadvantages observed here, three key concerns bear consideration. First, in a recent study of changes following the introduction of the Affordable Care Act (ACA), LGB individuals are now more likely to have insurance coverage than was true in the past, however they still experience barriers in access to and affordability of care (Skopec & Long, 2015). This is particularly salient in seeking and obtaining care from providers who are able to provide culturally competent care. Because much of the ACA coverage is new, those providers most competent to deliver such services may be in high demand or unavailable (Skopec & Long, 2015). Second, whether the practitioners LGB individuals do encounter will have the requisite cultural competence remains an open question (Hope & Chappell, 2015; Lyons et al., 2010; McGeorge & Stone Carlson, 2016). And finally, which approaches to counseling and psychotherapy are most effective at addressing stigma-related distress and needs for resiliency has yet to be determined (Pachankis, Hatzenbuehler, et al., 2015). Our ability to answer these questions or to track whether anti-LGB discrimination plays a role in access to effective MHSA-related care for LGB individuals is currently limited in existing national data sets. Nevertheless, it is essential that we ask and answer these questions as we evaluate the contribution of the ACA to addressing and eliminating sexual orientation-related health disparities. This is critical particularly in the area of MHSA-related functional limitations, where available workforce issues may play a significant role.

References


This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.


Hope, D. A., & Chappell, C. L. (2015). Extending training in multicultural competencies to include individuals identifying as lesbian, gay, and
bise
tical: Key choice points for clinical psychology training programs. 


Stata Corporation. (2013). Stata 13 user’s guide. College Station, TX: Stata Press.


Received May 10, 2016
Revision received July 12, 2017
Accepted July 13, 2017

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample

JOSHUA BRESLAU1*, SERGIO AGUILAR-GAXIOLA2, KENNETH S. KENDLER3, MAXWELL SU1, DAVID WILLIAMS4 AND RONALD C. KESSLER1

1 Department of Health Care Policy, Harvard Medical School, Boston, MA; 2 University of California, Davis School of Medicine, Sacramento, CA; 3 Departments of Psychiatry and Human Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA; 4 Department of Sociology and Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Background. Epidemiological studies have found lower than expected prevalence of psychiatric disorders among disadvantaged race-ethnic minority groups in the USA. Recent research shows that this is due entirely to reduced lifetime risk of disorders, as opposed to persistence. Specification of race-ethnic differences with respect to clinical and social characteristics can help identify the protective factors that lead to lower lifetime risk among disadvantaged minority groups.

Method. Data on 5424 Hispanics, non-Hispanic Blacks, and non-Hispanic Whites came from the National Comorbidity Survey Replication, a nationally representative survey conducted with the World Mental Health version of the Composite International Diagnostic Interview. Race-ethnic differences in risk of disorders were compared across specific diagnoses, ages of onset, cohorts and levels of education.

Results. Both minority groups had lower risk for common internalizing disorders: depression, generalized anxiety disorder, and social phobia. In addition, Hispanics had lower risk for dysthymia, oppositional-defiant disorder and attention deficit hyperactivity disorder; non-Hispanic Blacks had lower risk for panic disorder, substance use disorders and early-onset impulse control disorders. Lower risk among Hispanics, relative to non-Hispanic Whites, was found only among the younger cohort (age ≤43 years). Lower risk among minorities was more pronounced at lower levels of education.

Conclusion. The pattern of race-ethnic differences in risk for psychiatric disorders suggests the presence of protective factors that originate in childhood and have generalized effects on internalizing disorders. For Hispanics, but not for non-Hispanic Blacks, the influence of these protective factors has emerged only recently.

INTRODUCTION

Social adversity is commonly associated with increased risk for psychiatric disorders (Dohrenwend, 2000). However, community studies in the USA have not found elevated point prevalence of psychiatric disorders among disadvantaged racial and ethnic minority groups (Somervell et al. 1989; Kessler et al. 1994), despite higher levels of social adversity experienced by these groups (Clark et al. 1999; Williams, 1999; Turner & Lloyd, 2004). In the National Comorbidity Survey (NCS), the first nationally representative survey of psychiatric disorders conducted in the USA, we showed that the lower than expected point prevalence of DSM-III mental disorders among racial-ethnic minorities was attributable exclusively to differences in lifetime risk of these disorders.
as opposed to course of illness. Indeed, the course of illness tended to be more persistent among minorities than non-Hispanic Whites in that study (Breslau et al. 2005). Thus, understanding reasons for the lower risk of ever developing a mental disorder appears to be the key to understanding the lower than expected prevalence of mental disorders among disadvantaged racial and ethnic minority groups in the USA.

In this report we take the next step towards identifying causes of race-ethnic differences in lifetime risk of psychiatric disorders by specifying these differences in greater detail with respect to clinical and sociodemographic characteristics. Using data from a recent national survey, the National Comorbidity Survey Replication (NCS-R), we investigate variation in race-ethnic differences across (1) individual DSM-IV disorders, (2) age of onset of disorder, (3) birth cohorts, and (4) educational attainment. Results of these comparisons will help guide the investigation of potential protective factors that result in lower lifetime risk among race-ethnic minorities.

Our previous report considered race-ethnic differences in three broad classes of disorder: mood, anxiety and substance use disorders. However, it is important to know whether the differences we observed are consistent across disorders within these classes, implicating protective factors with generalized effects, or whether they are attributable to differences in a small number of disorders, implicating factors with disorder-specific effects (Aneshensel et al. 1991). In this analysis we test for variation in race-ethnic differences across individual disorders within each class of disorders. In addition, we examine an additional class of disorders, impulse control disorders, that was not included in the earlier survey.

Specifying race-ethnic differences in early versus late age of onset of disorder can help identify the developmental periods during which protective factors exert their influence. If, for instance, reduced risk of psychiatric disorder begins with early-onset disorders, we should consider protective factors present in childhood environments rather than adult social experiences. It has been suggested that stressors associated with minority race-ethnicity vary across the lifespan; minority children may be sheltered from the negative impact that discrimination in labor markets and other areas of adult life has on mental health in adults (Gore & Aseltine, 2003). This hypothesis predicts that the lower risk among minorities will be stronger for early-onset than for late-onset disorders.

Specification of race-ethnic differences with respect to birth cohorts can help identify causal factors by locating the differences historically. This is of interest because the social conditions affecting minorities in the USA have changed in recent decades in ways that influence the experience of social adversity, such as income (Levy, 1998) and educational attainment (Kao & Thompson, 2003). However, studies using distress scales have found that race-ethnic differences have not varied from the 1950s through the 1990s, despite these changes (Thomas & Hughes, 1986). We do not know the extent to which the same is true for lifetime risk of psychiatric disorders. Continuity in race-ethnic differences across birth cohorts would point to factors that are transmitted across generations despite historical changes in social conditions.

The association of disadvantaged minority status with lower lifetime risk of disorders suggests that the relationship between socioeconomic status (SES) and risk of onset may vary across race-ethnic groups (Williams et al. 1992; Williams, 1997). Two theories of such variation have been suggested. First, the ‘double jeopardy’ theory suggests that morbidity will be particularly elevated among low SES members of disadvantaged minority groups. This theory was supported by studies using distress scales that found higher levels of distress among non-Hispanic Blacks relative to non-Hispanic Whites only among those with low SES (Kessler & Neighbors, 1986; McLeod & Owens, 2004). Second, the ‘declining returns’ theory predicts the opposite pattern, with higher morbidity among minority groups at higher levels of SES. This theory, suggested by studies which found that minorities have lower economic returns to investment in educational credentials (Chiswick, 1988), has been supported in research on indicators of general physical health (Farmer & Ferraro, 2005). With respect to mental health the ‘declining returns’ pattern is also consistent with the suggestion that social stressors are most severe for middle class minorities who have the highest expectations but also face the
most severe competition in labor markets (Parker & Kleiner, 1966; Neckerman et al., 1999; Cole & Omari, 2003; Jackson & Stewart, 2003).

**METHOD**

**Sample**

As detailed elsewhere (Kessler et al., 2004b), the NCS-R is a nationally representative survey of English-speaking household residents aged 18 and older in the coterminous USA. Face-to-face interviews were carried out by professional interviewers from the Institute for Social Research at the University of Michigan between February 2001 and April 2003. The response rate was 70.9%. The survey was administered in two parts. Part I included a core diagnostic assessment of all respondents \((n=9282)\) that took an average of about one hour to administer. Part II included questions about risk factors, consequences, other correlates and additional disorders. In an effort to reduce respondent burden and control study costs, Part II was administered only to 5692 of the 9282 Part I respondents, including all Part I respondents with a lifetime disorder plus a probability subsample of other respondents \((n=5692)\). This paper examines Part II respondents who were Hispanic, non-Hispanic Black or non-Hispanic White \((n=5424)\). Interviewers explained the study and obtained verbal informed consent prior to beginning each interview. Recruitment and consent were approved by the Human Subjects Committees of Harvard Medical School and the University of Michigan.

The data were weighted to adjust for differential probabilities of selection, differential non-response, and residual differences between the sample and tract-level 2000 Census population on sociodemographic variables. An additional Part II weight adjusted for over-sampling of Part I cases. Weighting is described in more detail elsewhere (Kessler et al., 2004b).

**Measures**

**Diagnostic assessment**

NCS-R diagnoses are based on the World Mental Health Survey Initiative Version of the World Health Organization Composite International Diagnostic Interview (WMH-CIDI) (Kessler & Ustun, 2004) a fully structured lay-administered diagnostic interview that generates both ICD-10 (WHO, 1991) and DSM-IV (APA, 1994) diagnoses. Four classes of disorder, consisting of a total of 18 DSM-IV disorders, are examined here: anxiety disorders [panic disorder, agoraphobia without panic disorder, specific phobia, social phobia, generalized anxiety disorder (GAD), post-traumatic stress disorder (PTSD), obsessive-compulsive disorder (OCD)]; mood disorders (major depressive disorder, dysthymia, bipolar I and II disorders); a series of four disorders that share a common feature of difficulties with impulse-control [intermittent explosive disorder, oppositional-defiant disorder (ODD), conduct disorder, attention-deficit/hyperactivity disorder (ADHD)]; and two substance use disorders (alcohol abuse or dependence, drug abuse or dependence). As described elsewhere (Kessler et al., 2004a), blind clinical re-interviews with the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002) found generally good concordance with WMH-CIDI diagnoses for anxiety, mood, and substance use disorders. Impulse-control diagnoses have not been validated.

**Onset**

The timing of onset of disorders was determined retrospectively using reports of age of onset. These reports were obtained in the WMH-CIDI using a question series designed to avoid the implausible response patterns obtained using the standard CIDI age-of-onset question (Simon & VonKorff, 1995). The sequence began with a question designed to emphasize the importance of accurate response: ‘Can you remember your exact age the very first time you (HAD THE SYNDROME)?’ Respondents who answered ‘no’ were probed for a bound of uncertainty by moving up the age range incrementally (e.g. ‘Was it before you first started school?’, ‘Was it before you became a teenager?’, etc.). Age of onset was set at the upper end of the bound (e.g. age 12 for respondents who reported that onset was before they became teenagers). Experimental research shows this question sequence yields responses with a much more plausible age-of-onset distribution than the standard CIDI age-of-onset question (Knauper et al. 1999).
Race-ethnicity

Race-ethnicity was determined using the two-question format recommended by the U.S. Census Bureau on the basis of the Race and Ethnic Targeted Test (USA Census Bureau, 1997). Respondents were first asked if they were ‘of Hispanic or Latino descent’. All respondents were then asked ‘Which of the following best describes your race: American Indian, Alaska Native, Asian, Black or African-American, Native Hawaiian, Pacific Islander, or White?’ All respondents who indicated Hispanic or Latino descent are considered Hispanic. Non-Hispanics were then classified according to the race category they selected. Non-Hispanics who were neither Black nor White were excluded from this analysis.

Age of onset and cohort variables

Age of onset was dichotomized into early and late time periods using the median of the distribution of age of onset for each class of disorders. The cut-points were: 10 years of age for anxiety disorders, 22 years of age for mood disorders, 10 years of age for impulse control disorders and 18 years of age for substance use disorders. The sample was divided into younger (age ≤ 43) and older (age > 43) cohorts based on the median age of the sample.

Parental and respondent education

Data on parental and respondent education were used as indicators of SES in childhood and adulthood respectively. The educational attainment of the respondent’s parent (i.e. chief breadwinner in childhood) was classified into four categories (0–11, 12, 13–15, 16+). Respondent education was defined as a time-varying covariate assuming an orderly progression through years of education. Person-years were classified as (1) student prior to graduation from high school, (2) student after graduation from high school, (3) non-student with less than 12 years of education, (4) non-student with 12 years of education, (5) non-student with 13–15 years of education, (6) non-student with 16 or more years of education.

Statistical analysis

Lifetime prevalence was estimated as the proportion of respondents who ever had a given disorder up to their age at the time of the interview. Since race-ethnic groups differ in their age distributions, comparisons of lifetime prevalence are not equivalent to comparisons of lifetime risk. Lifetime risk was examined using discrete-time survival analysis with person-year the unit of analysis and statistical controls for age and sex (Efron, 1988). Variation in race-ethnic differences across disorders, cohorts, ages of onset, and education were evaluated by including interactions between these variables and race-ethnicity.

Standard errors of prevalence estimates and survival coefficients were estimated using the Taylor series linearization method (Wolter, 1985) implemented in the SUDAAN software system (Research Triangle Institute, 2002). Multivariate significance tests were made with Wald \( \chi^2 \) tests using Taylor series design-based coefficient variance–covariance matrices. All significance tests were evaluated at the 0.05 level with two-sided tests.

RESULTS

Sample characteristics

Reflecting the USA population, both Hispanics and non-Hispanic Blacks were younger (Meyer, 2001) and had lower levels of education (Bauman & Graf, 2003) and lower levels of parental education (Kao & Thompson, 2003) than non-Hispanic Whites (Table 1). Also consistent with USA census data, the ratio of men to women is highest among Hispanics and lowest among non-Hispanic Blacks (Spraggins, 2003).

Lifetime prevalence

Compared with non-Hispanic Whites, Hispanics had lower lifetime prevalence for any disorder, two of the four classes of disorders and six of the 16 specific disorders (Table 2). These differences were significant for dysthymia, depression, GAD and social phobia. For non-Hispanic Blacks, lower lifetime prevalence was more pervasive: non-Hispanic Blacks had lower prevalence than non-Hispanic Whites for any disorder, all four classes of disorders and for 13 of the 16 specific disorders. Differences were significant for any disorder, all classes of disorder except impulse control and for depression, GAD, panic disorder, social phobia and alcohol use disorder. There is only one
### Table 1. Sociodemographic characteristics of Hispanics, non-Hispanic Blacks and non-Hispanic Whites in the NCS-Ra

<table>
<thead>
<tr>
<th></th>
<th>Hispanic</th>
<th>Non-Hispanic Black</th>
<th>Non-Hispanic White</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>(S.E.)</td>
<td>%</td>
<td>(S.E.)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50.2</td>
<td>(5.0)</td>
<td>57.7</td>
<td>(1.9)*</td>
</tr>
<tr>
<td>Male</td>
<td>49.8</td>
<td>(5.0)</td>
<td>42.3</td>
<td>(1.9)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>37.4</td>
<td>(2.7)*</td>
<td>29.7</td>
<td>(2.8)*</td>
</tr>
<tr>
<td>30–44</td>
<td>35.4</td>
<td>(2.9)</td>
<td>30.5</td>
<td>(2.0)</td>
</tr>
<tr>
<td>45–59</td>
<td>18.0</td>
<td>(2.6)</td>
<td>23.9</td>
<td>(2.1)</td>
</tr>
<tr>
<td>60+</td>
<td>9.2</td>
<td>(2.3)</td>
<td>15.9</td>
<td>(2.3)</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>34.8</td>
<td>(3.7)*</td>
<td>22.0</td>
<td>(3.1)*</td>
</tr>
<tr>
<td>12</td>
<td>35.4</td>
<td>(2.9)</td>
<td>36.6</td>
<td>(4.7)</td>
</tr>
<tr>
<td>13–15</td>
<td>19.1</td>
<td>(2.4)</td>
<td>27.8</td>
<td>(1.6)</td>
</tr>
<tr>
<td>16+</td>
<td>10.8</td>
<td>(1.8)</td>
<td>13.6</td>
<td>(1.9)</td>
</tr>
<tr>
<td>Parental education (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>60.1</td>
<td>(3.7)*</td>
<td>40.5</td>
<td>(3.2)*</td>
</tr>
<tr>
<td>12</td>
<td>18.9</td>
<td>(2.4)</td>
<td>33.2</td>
<td>(2.2)</td>
</tr>
<tr>
<td>13–15</td>
<td>6.5</td>
<td>(0.7)</td>
<td>15.3</td>
<td>(2.5)</td>
</tr>
<tr>
<td>16+</td>
<td>14.5</td>
<td>(3.0)</td>
<td>11.0</td>
<td>(1.4)</td>
</tr>
<tr>
<td>(n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(527)</td>
<td>(717)</td>
<td>(4180)</td>
<td>(5424)</td>
</tr>
</tbody>
</table>

*a Part II NCS-R respondents who are either Hispanic, non-Hispanic Black, or non-Hispanic White. All respondents in other race-ethnic groups (n = 268) are excluded.

* Significant difference from non-Hispanic Whites evaluated with $\chi^2$ test at $p = 0.05$.

### Table 2. Comparison of lifetime prevalence of psychiatric disorders across race-ethnic groups

<table>
<thead>
<tr>
<th></th>
<th>Hispanic</th>
<th>Non-Hispanic Black</th>
<th>Non-Hispanic White</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>(S.E.)</td>
<td>%</td>
<td>(S.E.)</td>
</tr>
<tr>
<td>I. Anxiety disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>5.4</td>
<td>(0.9)</td>
<td>3.1*</td>
<td>(0.6)</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>4.8*</td>
<td>(0.9)</td>
<td>5.1*</td>
<td>(0.7)</td>
</tr>
<tr>
<td>Agoraphobia without panic</td>
<td>2.7</td>
<td>(0.6)</td>
<td>2.3</td>
<td>(0.5)</td>
</tr>
<tr>
<td>Social phobia</td>
<td>8.8*</td>
<td>(1.3)</td>
<td>10.8</td>
<td>(1.2)</td>
</tr>
<tr>
<td>Specific phobia</td>
<td>13.1</td>
<td>(1.5)</td>
<td>11.7</td>
<td>(1.2)</td>
</tr>
<tr>
<td>Post-traumatic stress disorder</td>
<td>5.9</td>
<td>(1.0)</td>
<td>7.1</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>1.2</td>
<td>(0.6)</td>
<td>0.5</td>
<td>(0.2)</td>
</tr>
<tr>
<td>Any anxiety</td>
<td>24.9</td>
<td>(2.2)</td>
<td>23.8*</td>
<td>(1.9)</td>
</tr>
<tr>
<td>II. Mood disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depression</td>
<td>13.5*</td>
<td>(1.5)</td>
<td>10.8*</td>
<td>(1.2)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>4.3</td>
<td>(1.0)</td>
<td>4.9*</td>
<td>(0.6)</td>
</tr>
<tr>
<td>Any</td>
<td>17.9</td>
<td>(1.8)</td>
<td>14.5</td>
<td>(1.9)</td>
</tr>
<tr>
<td>III. Impulse control disorders</td>
<td>9.9</td>
<td>(1.4)</td>
<td>7.2</td>
<td>(1.1)</td>
</tr>
<tr>
<td>Conduct</td>
<td>6.9</td>
<td>(1.0)</td>
<td>4.9</td>
<td>(1.2)</td>
</tr>
<tr>
<td>Any</td>
<td>17.9</td>
<td>(1.8)</td>
<td>14.5</td>
<td>(1.9)</td>
</tr>
<tr>
<td>IV. Substance disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol abuse or dependence</td>
<td>15.0</td>
<td>(2.3)</td>
<td>9.5*</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Drug abuse or dependence</td>
<td>9.1</td>
<td>(1.4)</td>
<td>6.3</td>
<td>(1.0)</td>
</tr>
<tr>
<td>Any</td>
<td>16.1</td>
<td>(2.3)</td>
<td>10.8*</td>
<td>(1.0)</td>
</tr>
<tr>
<td>V. Any disorder</td>
<td>43.7</td>
<td>(4.0)</td>
<td>38.5*</td>
<td>(2.8)</td>
</tr>
</tbody>
</table>

* Significant difference from non-Hispanic Whites evaluated with $\chi^2$ test at $p = 0.05$. 

**Race-ethnic differences in risk of psychiatric disorders**
instance in which a minority group had significantly higher lifetime prevalence relative to non-Hispanic Whites: non-Hispanic Blacks had significantly higher lifetime prevalence of bipolar disorder.

**Lifetime risk**

Comparisons of lifetime risk, estimated in survival models that take into account differences in the age distribution across race-ethnic groups, show an even stronger pattern of lower risk for disorders among minorities than was indicated by comparisons of lifetime prevalence (Table 3). Both Hispanics and non-Hispanic Blacks had significantly lower lifetime risk for mood disorders, anxiety disorders and for any psychiatric disorder, and non-Hispanic Blacks had lower lifetime risk for substance use disorders. Variation in race-ethnic differences across the individual disorders within a class was significant in all but three cases: impulse control disorders among non-Hispanic Blacks, and substance use disorders among both minority groups. Both groups had lower lifetime risk than non-Hispanic Whites for depression, GAD and social phobia. There were also group specific differences: compared to non-Hispanic Whites, Hispanics had lower lifetime risk for dysthymia, oppositional defiant disorder, and ADHD, while non-Hispanic Blacks had lower lifetime risk for substance use disorders and panic disorder.

**Variations across age of onset**

Race-ethnic differences did not vary significantly between early and late onset disorders ($\chi^2_{1}=0.1–2.4, p=0.800–0.121$), with one exception. Non-Hispanic Blacks had significantly lower risk for early-onset (prior to age 10) impulse control disorders, but not for late-onset

<table>
<thead>
<tr>
<th>Table 3. Race-ethnic differences in lifetime risk of DSM-IV disorders</th>
<th>Hispanic</th>
<th>Non-Hispanic Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>I. Anxiety disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any anxiety disorder</td>
<td>0.8* (0.6–0.9)</td>
<td>0.7* (0.6–0.9)</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>1 (0.8–1.4)</td>
<td>0.6* (0.4–0.8)</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>0.6* (0.4–0.8)</td>
<td>0.6* (0.4–0.8)</td>
</tr>
<tr>
<td>Agoraphobia without panic</td>
<td>1.1 (0.6–1.9)</td>
<td>0.9 (0.6–1.5)</td>
</tr>
<tr>
<td>Social phobia</td>
<td>0.6* (0.4–0.9)</td>
<td>0.8* (0.6–1.0)</td>
</tr>
<tr>
<td>Specific phobia</td>
<td>1 (0.8–1.4)</td>
<td>0.9 (0.7–1.1)</td>
</tr>
<tr>
<td>Post-traumatic stress disorder</td>
<td>0.9 (0.6–1.2)</td>
<td>1 (0.7–1.3)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>2.7 (0.8–8.9)</td>
<td>1 (0.4–2.5)</td>
</tr>
<tr>
<td>II. Mood disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any mood disorder</td>
<td>0.7* (0.6–0.9)</td>
<td>0.6* (0.5–0.8)</td>
</tr>
<tr>
<td>Dysthymic disorder</td>
<td>0.5* (0.3–0.8)</td>
<td>0.8 (0.6–1.1)</td>
</tr>
<tr>
<td>Major depression</td>
<td>0.7* (0.5–0.9)</td>
<td>0.6* (0.4–0.7)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>1 (0.7–1.6)</td>
<td>1.2 (0.9–1.5)</td>
</tr>
<tr>
<td>III. Impulse disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any impulse disorder</td>
<td>0.8 (0.6–1.1)</td>
<td>0.8 (0.6–1.1)</td>
</tr>
<tr>
<td>Intermittent explosive disorder</td>
<td>1 (0.7–1.5)</td>
<td>—</td>
</tr>
<tr>
<td>Conduct disorder</td>
<td>0.9 (0.7–1.3)</td>
<td>—</td>
</tr>
<tr>
<td>Oppositional defiant disorder</td>
<td>0.6* (0.4–0.9)</td>
<td>—</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>0.6* (0.4–1.0)</td>
<td>—</td>
</tr>
<tr>
<td>IV. Substance disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any substance disorder</td>
<td>1 (0.7–1.3)</td>
<td>0.7* (0.6–0.9)</td>
</tr>
<tr>
<td></td>
<td>$\chi^2_{1}=0.2 (p=0.642)$</td>
<td>$\chi^2_{1}=1.5 (p=0.217)$</td>
</tr>
<tr>
<td>V. Any disorder</td>
<td>0.8* (0.6–1.0)</td>
<td>0.7* (0.6–0.8)</td>
</tr>
</tbody>
</table>

OR, Odds ratio; CI, confidence interval.

* OR estimated in discrete time survival models controlling for sex and age. Non-Hispanic Whites are the reference group. $\chi^2$ represent tests for variation in difference (OR) between minority group and non-Hispanic Whites across individual disorders within each class of disorders. * Significant difference between minority group and non-Hispanic Whites at $p=0.05$ level, two-sided test.
impulse control disorders \(\chi^2_1 = 4.5, p = 0.034\): OR (early-onset) 0.7, 95% CI 0.5–1.0; OR (late onset) 1.0, 95% CI 0.7–1.5].

Variations across cohorts

Analysis of cohort variation in race-ethnic differences revealed that the lower lifetime risk among Hispanics is attributable entirely to the younger cohort, i.e. those aged 43 or younger (Table 4). Significant variation was found for each of the four classes of disorder (\(\chi^2_1 = 4.3–7.2, p = 0.039–0.007\)). Specifically, in the older cohort, there were no significant differences in risk for anxiety, mood and impulse control disorders between Hispanics and non-Hispanic Whites, but in the younger cohort Hispanics had significantly lower risk than non-Hispanic Whites for each of these classes of disorder. Results for substance use disorders showed a trend in the same direction: in the older cohort Hispanics had significantly higher lifetime risk, whereas in the younger cohort there was no significant difference between Hispanics and non-Hispanic Whites.

Variations across parental and respondent education

Lower lifetime risk among both minority groups was most pronounced at lower levels of education (Table 5). For Hispanics, there was significant variation in the odds ratio for three of the four classes of disorder: parental education for substance use disorders (\(\chi^2_3 = 0.6, p = 0.022\)) and impulse control disorders (\(\chi^2_3 = 9.3, p = 0.026\)) and respondent’s education for mood disorders (\(\chi^2_5 = 18.6, p = 0.002\)). In each instance, lower lifetime risk in Hispanics relative to non-Hispanic Whites was restricted to lower education levels. For non-Hispanic Blacks there was significant variation across respondent education for impulse control disorders (\(\chi^2_5 = 21.3, p = 0.001\)) with the same trend as observed in Hispanics. No other interactions were detected for parental education (\(\chi^2_5 = 1.7–6.3, p = 0.645–0.098\)) or respondent education (\(\chi^2_5 = 2.0–10.4, p = 0.849–0.064\)).

DISCUSSION

Consistent with previous national surveys in the USA, data from the NCS-R indicate that Hispanics and non-Hispanic Blacks have lower lifetime risk of psychiatric disorders than non-Hispanic Whites (Breslau et al. 2005). This finding suggests the need to investigate potential protective factors that might explain the advantage that race-ethnic minority groups enjoy despite the social disadvantages they experience. Ethnic identification (Herd & Grube, 1996; Mossakowski, 2003) and religious participation (Wallace & Forman, 1998; Varon & Riley, 1999; Ellison et al. 2001; Lee & Newberg, 2002).
Lower lifetime risk among both Hispanics and non-Hispanic Blacks relative to non-Hispanic Whites begins in childhood (i.e. prior to age 10).

(3) Differences in lifetime risk favoring minorities are more pervasive in the younger cohort (i.e. born circa 1958 or more recently).

(4) Variation in race-ethnic differences across levels of education is sporadic and, where it occurs, it shows that lower lifetime risk in members of minority groups is more pronounced at lower levels of education (i.e. parent or respondent had less than high school education).

Lower risk for mood and anxiety disorders among both minority groups was accounted for by five disorders: major depression, dysthymia, GAD, social phobia, and panic disorder. These disorders fall within a set of ‘internalizing’ disorders, which cluster together in factor analyses, suggesting that they may share etiologic pathways (Krueger, 1999; Kendler et al. 2003). The finding that these disorders vary together in lifetime risk across race-ethnic groups suggests the presence of common protective factors across disorders for both minority groups. This possibility would be further supported if future analyses show that patterns of co-morbidity among these disorders are similar across race-ethnic groups. Substance use and impulse control disorders, which comprise the externalizing factor identified in analyses of co-morbidity, do not differ consistently across race-ethnic groups, suggesting that the explanation for the observed differences in these disorders is more complex.

The finding that race-ethnic differences emerge in childhood strongly suggests that the search for causes should focus on factors present in the early life environment. A focus on childhood is a particularly important consideration for studies of religious participation and ethnic identification. These factors are measured in adults, but may also reflect the cultural contexts in which children are socialized, rather than the direct effect of religious participation on adult experience.

We found that for Hispanics, differences in lifetime risk relative to non-Hispanic Whites varied significantly across birth cohorts for all classes of disorder, and that the phenomenon

<table>
<thead>
<tr>
<th>Table 5. Variations in race-ethnic differences in lifetime risk across respondent education and parental education(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Mood disorders</strong></td>
</tr>
<tr>
<td>Respondent education(^b)</td>
</tr>
<tr>
<td>Student &lt; HS</td>
</tr>
<tr>
<td>Student &gt; HS</td>
</tr>
<tr>
<td>Less than HS</td>
</tr>
<tr>
<td>HS graduate</td>
</tr>
<tr>
<td>13–15 years</td>
</tr>
<tr>
<td>16+ years</td>
</tr>
</tbody>
</table>

| **II. Substance use disorders** |
| Parental education |
| < HS | 0.8 (0.5–1.2) |
| HS | 1.1 (0.5–2.1) |
| 13–15 years | 2.3 (0.9–6.0) |
| 16+ years | 1.1 (0.6–2.3) |

| **III. Impulse control disorders** |
| Parental education |
| < HS | 0.5 (0.3–0.7) |
| HS | 0.8 (0.4–1.6) |
| 13–15 years | 1.2 (0.6–2.4) |
| 16+ years | 1.3 (0.8–2.3) |

| Respondent education | Blacks v. Whites |
| Student < HS | 0.7 (0.3–1.4) |
| Student > HS | 2.7 (1.2–6.0) |
| Less than HS | 0.5 (0.2–1.6) |
| HS graduate | 0.8 (0.3–2.3) |
| 13–15 years | 2.1 (0.7–6.4) |
| 16+ years | 3.6 (0.8–15.3) |

OR, Odds ratio; CI, confidence interval; HS, High School.
\(^a\) Strata-specific OR shown where interactions between education and race-ethnicity were significant. See text for significance tests. Estimates derived from discrete time survival models with controls for age and sex.
\(^b\) Respondent education is coded as a time-varying covariate as explained in the text.

2005) have been suggested as protective factors that might explain lower lifetime risk for disorder among race-ethnic minorities. However, cultural and social factors that might be related to risk for psychiatric disorder are complex and have undergone significant changes in recent decades.

As noted in the introduction, our goal in this analysis was to advance the examination of suspected protective factors by investigating variation in race-ethnic differences across specific disorders, age of onset of disorders and subgroups of the population defined by cohort and education. The key findings of this analysis are:

(1) Lower lifetime risk among race-ethnic minority groups relative to non-Hispanic Whites is attributable to a small number of common disorders.
of lower lifetime risk of disorders was specific to the younger cohort. For non-Hispanic Blacks, a similar trend was found for substance use disorders. It is unclear whether the protective factors suggested in the literature, ethnic identification and religious participation, have followed a similar historical trajectory. There is no evidence that relative rates of religious participation across groups have changed, but there is evidence that religious affiliations within both minority groups have shifted toward conservative protestant groups that tend to promote abstention from alcohol (Hunt, 1999; Sherkat, 2002). The role of religion could be further elucidated if future studies examine whether observed historical changes in race-ethnic differences in risk of disorder are related to contemporaneous changes in specific aspects of religiosity. Future studies should also examine historical changes in patterns of immigration (Portes & Zhou, 2003), educational achievement (Kao & Thompson, 2003) and income (Levy, 1998) across minority groups.

Analysis of variation in race-ethnic differences across levels of SES, using parental and respondent education as indicators, contradict the ‘double jeopardy’ theory, which would predict that low SES minorities would have the highest levels of risk because of the combined effect of socio-economic and race-ethnic disadvantage. Where variation in race-ethnic differences in lifetime risk was found, the pattern was consistent with predictions of the ‘declining returns’ theory, which predicts that minority groups enjoy fewer of the health improvements that come with higher SES. This pattern is consistent with research on Hispanics which found that Hispanic enclave communities offer social resources to their members that may have a salutary impact on health (Sanders, 2002; Portes & Zhou, 2003; Eschbach et al. 2004).

Limitations
The findings should be interpreted in light of several limitations. First, the assessment of lifetime risk for psychiatric disorders depends on the accuracy of lifetime recall of psychiatric symptoms. As mentioned in the Method section above, the WMH-CIDI included question sequences found in experimental research to improve recall of life events in survey research. There is no evidence that recall of psychiatric symptoms is differential across race-ethnic groups.

Second, differential non-response across race-ethnic groups may have led to underestimation of lifetime disorders among minority groups (USA DHHS, 2001). To minimize the impact of non-response bias in this survey, a non-respondent survey was carried out in which a representative subsample of initial non-respondents was administered a brief telephone survey that collected screening information about anxiety, mood, impulse-control, and substance disorders (Kessler et al. 2004b). Results were used to weight the main survey data to adjust for any differential under-representation of disorders on the basis of race-ethnicity as well as other sociodemographic factors. In addition, missing data on parental education, a problem that was more common for minorities than for non-Hispanic Whites, were imputed based on information on parental occupation. Results did not change meaningfully when observations with missing values were removed from the analysis, but it remains possible that residual bias remains that might limit the generalizability of the results reported here.

Third, cultural differences may lead members of different race-ethnic groups to respond differently to the same survey questions regarding their psychiatric history despite similar levels of morbidity (Rogler, 1999; Rogler et al. 2001). Psychometric studies of differential item functioning examining distress scales have found some differences along these lines across race-ethnic groups. However, these differences generally overestimated levels of distress of minority groups and were attributable to items related to positive mental health (Iwata & Buka, 2002). Positive mental health items are not used to assess disorders in this study. Reporting of psychiatric symptoms might also be influenced by the race-ethnicity of the interviewers. Since the finding of lower lifetime risk among minority groups is consistent across studies using different structured diagnostic instruments, e.g. the Diagnostic Interview Schedule (DIS) (Somervell et al. 1989; Weissman et al. 1991), the University of Michigan (UM)-CIDI (Kessler et al. 1994) and the WMH-CIDI (Kessler et al. 2005), methodological studies of differential validity across race-ethnic groups...
should focus on methods shared by these instruments.

We have focused on race-ethnic differences in lifetime risk because of the counterintuitive finding of lower lifetime risk among disadvantaged groups. However, this finding of lower lifetime risk does not imply that disadvantaged groups suffer a lower overall burden of psychiatric morbidity. Our previous work suggests that the advantage of reduced lifetime risk is offset by the disadvantage of an increased risk for persistence among members of minority groups who become ill (Breslau et al. 2005). Assessment of the aggregate public health implications of race-ethnic differences in psychiatric morbidity, therefore, will require comparisons that consider multiple aspects of the course of disorders, including persistence as well as severity and disability.

ACKNOWLEDGEMENTS

The National Comorbidity Survey was supported by the National Institute of Mental Health (NIMH; R01 MH46376, R01 MH49098, and R01 MH52861) with supplemental support from the National Institute of Drug Abuse (NIDA; through a supplement to MH46376) and the W. T. Grant Foundation (90135190). The National Comorbidity Survey Replication (NCS-R) is supported by NIMH (U01-MH60220) with supplemental support from NIDA, the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044708), and the John W. Alden Trust.

Collaborating investigators include Ronald C. Kessler (Principal Investigator, Harvard Medical School), Kathleen Merikangas (Co-Principal Investigator, NIMH), James Anthony (Michigan State University), William Eaton (The Johns Hopkins University), Meyer Glantz (NIDA), Doreen Koretz (Harvard University), Jane McLeod (Indiana University), Mark Olfson (Columbia University College of Physicians and Surgeons), Harold Pincus (University of Pittsburgh), Greg Simon (Group Health Cooperative), Michael Von Korff (Group Health Cooperative), Philip Wang (Harvard Medical School), Kenneth Wells (UCLA), Elaine Weithington (Cornell University), and Hans-Ulrich Wittchen (Max Planck Institute of Psychiatry). The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or the USA Government. A complete list of NCS publications and the full text of all NCS-R instruments can be found at http://www.hcp.med.harvard.edu/ncs. Send correspondence to NCS@hcp.med.harvard.edu. The NCS-R is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative.

We thank the staff of the WMH Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis. These activities were supported by the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the USA Public Health Service (R13-MH066849, R01-MH069864, and R01 DA016558), the Pan American Health Organization, Eli Lilly and Company, and GlaxoSmithKline. A complete list of WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/. Support for this analysis was provided by NIMH (K01 MH66057-02).

DECLARATION OF INTEREST

None.

REFERENCES


Race-ethnic differences in risk of psychiatric disorders


USA Census Bureau (1997). Results of the 1996 Race and Ethnic Targeted Test. USA Census Bureau: Washington, DC.


A Theory of Unmitigated Communion

Vicki S. Helgeson and Heidi L. Fritz
Department of Psychology
Carnegie Mellon University

Research has established that women suffer more often than men from depression. Sex role socialization has been offered as one explanation for this sex difference, but traditional measures of female gender-related traits are not related to depressive symptoms. We argue that thus far research has failed to distinguish the traditional measure of female gender-related traits, communion, from another set of gender-related traits, unmitigated communion. Unmitigated communion is a focus on and involvement with others to the exclusion of the self. Unmitigated communion, but not communion, is related to psychological distress, including depressive symptoms, and accounts for sex differences in distress. We examine the relation of unmitigated communion to communion as well as other personality constructs and then describe the cognitive and behavioral features of unmitigated communion. We note the implications of unmitigated communion for physical and psychological well-being and speculate on possible origins.

From adolescence onward, two times more American women than men report experiencing depression (Nolen-Hoeksema, 1987). A common explanation for this gender difference is that women are socialized to be more concerned with others than themselves (Miller, 1976). However, traditional measures of female gender-related traits do not correlate with depressive symptoms, and thus cannot explain women's comparatively greater depression than men. In addition, the mechanisms by which women's focus on others leads to distress are not well specified. An explanation for this puzzle may lie in the fact that past research has failed to distinguish between two sets of female gender-related traits, only one of which is linked to distress. We argue that sex-role socialization may lead to the development of two very different gender-related traits: communion and unmitigated communion. The former accounts for women's closer relationships and superior interpersonal skills compared to men; the latter accounts for women's greater psychological distress compared to men. Unmitigated communion is characterized by an excessive concern with others and placing others' needs before one's own. We distinguish unmitigated communion from other personality constructs, describe its cognitive and behavioral manifestations, discuss its implications for well-being, and speculate on possible origins.

Communion is a broad personality construct that Bakan (1966) first introduced as a fundamental modality of human existence. He described communion as a focus on connection with others and a focus that was more likely to characterize women than men in our culture. As a specific personality construct, communion has been assessed with instruments initially referred to as measures of psychological femininity (Bem, 1974; Spence, Helmreich, & Stapp, 1974), but are now thought to reflect one aspect of female gender-related traits, a communal or expressive orientation (Spence, 1984). These instruments certainly do not capture the entire communal modality of human existence as articulated by Bakan, but they do tap traits consistent with a communal orientation. These scales reflect a focus on others and relationships and on making and maintaining connections with others. Items on these instruments, specifically the Personal Attributes Questionnaire (PAQ; Spence et al., 1974) and the Bem Sex Role Inventory (BSRI; Bem, 1974), include "kind," "warm in relations," "helpful to others," "gentle," and "understanding." The PAQ and BSRI communion scales are highly correlated with the warm/agreeable dimension of the Interpersonal Adjective Scales (Wiggins & Broughton, 1985). These communion scales, however, are not related to distress, and thus cannot account for
women's greater distress than men (see Helgeson, 1994, for a review).

Yet, socialization pressures to focus on others may still account for women's greater distress compared to men. We outline a specific kind of other-focus that may be responsible for women's distress: unmitigated communion. Unmitigated communion is a focus on others to the exclusion of the self, resulting in the neglect of one's own needs. Although Bakan (1966) never specifically identified the construct of unmitigated communion, he discussed the idea that high levels of communion ought to be mitigated by a personal sense of agency (i.e., focus on self and separation). He also has agreed with recent discussions of the conceptualization of unmitigated communion (Bakan, personal communication, November 3, 1995). Unmitigated communion is a form of communion in which agency is notably absent (Helgeson & Fritz, 1997). In fact, high communion and low agency may predispose one to develop unmitigated communion.

What is unmitigated communion? We broadly conceptualize unmitigated communion as a focus on others to the exclusion of the self. We have operationalized unmitigated communion in the form of a self-report scale (Fritz & Helgeson, in press; Helgeson, 1993) that reflects placing others' needs before one's own (e.g., "I always place the needs of others above my own"; "I can't say no when someone asks me for help") and an excessive concern with others (e.g., "I worry about how other people get along without me when I am not there"; "I have no trouble sleeping at night when other people are upset" [latter is reverse scored]). The nine-item scale is shown in the Appendix. A person who scores high on unmitigated communion would seem to have high levels of communion and low levels of agency (see Helgeson & Fritz, 1997, for a review), but high communion and low agency is not to be equated with unmitigated communion. A high focus on others and a low focus on the self does not necessarily imply that one has an excessive concern with others and that one places others needs before one's own. A high focus on others and a low focus on the self also does not necessarily lead to the cognitive and behavioral manifestations of unmitigated communion that we outline herein (i.e., low self-esteem, becoming overly involved with others, self-neglect). We argue at the end of this article that certain environmental events may lead someone with a communal orientation to develop unmitigated communion.

In this article, we first describe the relation of unmitigated communion to major personality constructs in the literature to demonstrate that we are identifying a unique personality trait. We refer to unmitigated communion as a personality trait because it is relatively stable over time (Helgeson, 1993). We do not, however, assume unmitigated communion is genetically based. In fact, we believe unmitigated communion is most likely to be rooted in environmental events and that one is capable of changing this style of interacting with others to some degree.

Next, we describe the cognitive and behavioral features of unmitigated communion, paying particular attention to those features that distinguish it from communion. Third, we discuss the implications of unmitigated communion for psychological and physical well-being. Finally, we speculate on some possible origins of unmitigated communion. The reader should assume the data discussed in this article are correlational unless otherwise noted. Throughout this review, we rely mostly on previously published articles. However, at times we collected data purposely for this article, and we present those data here. Data that have not been previously published are presented herein along with appropriate statistical tests. Findings that have been reported elsewhere are not accompanied by statistical tests; instead, we direct the reader to the respective reports of the findings.

Relation of Unmitigated Communion to Personality Constructs

Unmitigated communion is positively correlated with communion, yet the two reveal divergent relations to well-being (Fritz & Helgeson, in press; see Helgeson, 1994, for a review). We examined the relation of unmitigated communion, as well as communion, to the Big Five personality traits (Goldberg, 1992) in a sample of 526 undergraduates (318 male, 208 female). We used abbreviated eight-item measures of the Big Five, previously employed by Cohen, Doyle, Skoner, Rabin, and Gwaltney (1997). The internal consistencies of these scales ranged from .77 to .90.

Unmitigated communion was clearly not reducible to the Big Five. Its highest correlation was with agreeableness ($r = 36$, $p < .001$), but communion was more strongly associated with agreeableness ($r = .66$, $p < .001$). In addition, when communion was statistically controlled, the correlation between unmitigated communion and agreeableness disappeared ($r = .03$, $n.s.$). Thus, the relation of unmitigated communion to agreeableness is entirely accounted for by communion. Unmitigated communion also revealed a small negative correlation with emotional stability ($r = -17$, $p < .001$), which is consistent with previous research (Helgeson, 1993) and is not surprising because traits comprising the emotional stability scale include anxiety and depression items (e.g., anxious, nervous, sad, depressed). Unmitigated communion also revealed a small positive correlation with extraversion ($r = .10$, $p < .05$), but communion was more strongly associated with extraversion ($r = .25$, $p < .001$). Again, when communion was statistically controlled, the relation of unmitigated communion to extraversion disappeared ($r = -03$, $n.s.$). Unmitigated communion was not associated with con-
scientiousness or intellect, but communion was associated with both \( r = .13, \ p < .005; \ r = .17, \ p < .001 \), respectively.

To provide further evidence that unmitigated communion is not captured by the Big Five, we entered all five personality traits in a regression analysis to predict unmitigated communion. The Big Five explained 24\% of the variance in unmitigated communion compared to 54\% of the variance in communion. Thus, communion, but not unmitigated communion, can be more clearly mapped onto the Big Five personality traits.

Because unmitigated communion is a relationship-oriented construct, we also examined its association to adult attachment styles. To the extent that adult attachment styles are rooted in infancy (a subject of great debate), they could contribute to the development of unmitigated communion. Unmitigated communion is associated with insecure attachment styles, specifically the anxious/ambivalent style (also called preoccupied), whereas communion is more likely to be associated with the secure attachment style. Sample items from the anxious/ambivalent style are “Others often are reluctant to get as close as I would like,” and “I often worry that my partner(s) doesn’t (don’t) really love me.” Sample items from the secure attachment style are “I rarely worry about being abandoned by others,” and “I find it relatively easy to get close to others.” In an unpublished study of 284 college students, multi-item scales of Hazan and Shaver’s (1987) three attachment styles (secure, anxious/ambivalent, avoidant; Simpson, Rholes, & Nelligan, 1992) were administered. In a regression analysis, communion predicted higher scores on the secure attachment dimension \((\beta = .43, \ p < .001)\), whereas unmitigated communion marginally predicted lower scores on the secure attachment dimension \((\beta = -11, \ p = .08)\). Communion was associated with lower anxious/ambivalent scores \((\beta = -21, \ p < .01)\), whereas unmitigated communion was associated with higher scores \((\beta = 23, \ p < .01)\).

In a recent study of precollege students, one item measures of Bartholomew and Horowitz’s (1991) four attachment styles were administered. Unmitigated communion was positively associated with both of the insecure attachment styles (preoccupied and fearful), whereas communion was positively related to the secure and preoccupied styles and inversely related to the dismissing style (Fritz & Helgeson, in press, Study 1).

Thus, despite the positive correlation between unmitigated communion and communion, the two are associated with different sets of personality constructs. Unmitigated communion shows modest associations with agreeableness and extraversion that are accounted for entirely by its relation to communion. Unlike communion, unmitigated communion reveals modest associations with emotional instability and insecure adult attachment styles. Communion, however, is strongly associated with high agreeableness and reveals modest associations with extroversion, conscientiousness, intellect, and the secure adult attachment style.

Cognitions About Self and Others

Despite the positive correlation between communion and unmitigated communion, the two constructs differ along a number of dimensions. As described later, unmitigated communion is associated with a negative view of the self, and low self-esteem is one reason that unmitigated communion individuals become overly involved with others and neglect the self (Fritz & Helgeson, in press). Unmitigated communion also is associated with external standards for self-evaluation (Fritz & Helgeson, in press), meaning that people high in unmitigated communion depend on others for self-esteem. Thus, the unmitigated communion individual’s orientation toward others is partly based on a concern with others’ views of the self. By contrast, communion reflects a positive orientation toward others, has no implications for views of the self, and is not associated with overinvolvement with others or the neglect of the self. The differential associations of unmitigated communion and communion to views of self and views of others are shown in Table 1 and discussed later.

View of Self

Unmitigated communion is associated with a low regard for the self using both the Rosenberg (1965) Self-Esteem scale and Feyer’s (1953) Self Acceptance scale (Fritz & Helgeson, in press, Studies 1 and 3). Communion is unrelated to these measures of self-regard (Fritz & Helgeson, in press, Studies 1 and 3; see Helgeson, 1994, for a review).

Others’ View of Self

One reason people high in unmitigated communion have low self-esteem is that they base their self-esteem on others’ perceptions of themselves and perceive that others hold a negative view of the self. Unmitigated communion, but not communion, is associated with an externalized self-perception, meaning such individuals have external standards for self-evaluation (Fritz & Helgeson, in press, Studies 1 and 4). Unmitigated communion also is associated with a high fear of negative evaluation by others (Helgeson, Richardson, & Fritz, 1997) and a perception that others do not regard the self favorably (Fritz & Helgeson, in press, Study 1; Helgeson et al., 1997). The combination of basing one’s self-perception on others and perceiving that others do not value the self accounts for the relation of unmitigated communion to low self-esteem (Fritz & Hel-
Table 1. Commonalities and Distinctions Between Unmitigated Communion and Communion

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unmitigated Communion</th>
<th>Communion</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>View of Self and Others</td>
<td>—</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press); see Helgeson (1994) for a review</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Externalized Self-Perception</td>
<td>+</td>
<td>O</td>
<td>Helgeson, Richardson, &amp; Fritz (1997)</td>
</tr>
<tr>
<td>Fear of Negative Evaluation</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press); Helgeson et al. (1997)</td>
</tr>
<tr>
<td>Others’ Negative View of Self</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press); Helgeson et al. (1997)</td>
</tr>
<tr>
<td>Positive View of Others</td>
<td>O</td>
<td>+</td>
<td>Fritz &amp; Helgeson (in press); Helgeson et al. (1997)</td>
</tr>
<tr>
<td>Interpersonal Behaviors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide Social Support</td>
<td>++</td>
<td>+</td>
<td>Helgeson (1993); Helgeson &amp; Fritz (1995)</td>
</tr>
<tr>
<td>Overprotective Behavior</td>
<td>+</td>
<td>O</td>
<td>Helgeson (1993)</td>
</tr>
<tr>
<td>Overly Nurturing</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Intrusive</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Self-Sacrificing</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Uncomfortable Receiving Social Support</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Difficulty With Self-Disclosure</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Feel Bad When Others Ignore One’s Advice</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Empathy Axiomism Hypothesis</td>
<td>O</td>
<td>+</td>
<td>Helgeson et al. (1997)</td>
</tr>
<tr>
<td>Empathy Reward/Punishment Hypothesis</td>
<td>+</td>
<td>O</td>
<td>Helgeson et al. (1997)</td>
</tr>
<tr>
<td>Difficulty With Assertiveness</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Easily Exploitable</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Inhibit Self-Expression</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negatively Affected by Relationship Stressors</td>
<td>+</td>
<td>O</td>
<td>Helgeson &amp; Fritz (1996)</td>
</tr>
<tr>
<td>Intrusive Thoughts About Others’ Problems</td>
<td>+</td>
<td>O</td>
<td>Fritz &amp; Helgeson (in press)</td>
</tr>
<tr>
<td>Poor Adjustment to Disease</td>
<td>+</td>
<td>O</td>
<td>Helgeson (1993, 1997); Helgeson &amp; Fritz (1996)</td>
</tr>
<tr>
<td>Poor Metabolic Control</td>
<td>+</td>
<td>O</td>
<td>Helgeson &amp; Fritz (1996)</td>
</tr>
</tbody>
</table>

Note: — = negative correlation, + = positive correlation, O = no correlation; ++ compared to + indicates a stronger positive correlation for one construct than the other.
geson, in press, Study 1). Thus, one reason the other-focus of unmitigated communion has adverse consequences is because one’s self-esteem depends on others and one perceives others as evaluating the self negatively.

View of Others

Communion and empathy are clearly associated with the belief that other people are good and valuable, whereas unmitigated communion is not related to a positive or negative view of others (Fritz & Helgeson, in press, Study 1; Helgeson et al., 1997). People who score high on unmitigated communion probably have a mixed view of others. They place others’ needs before their own, suggesting a positive orientation toward other people. However, individuals characterized by unmitigated communion also may have a history of problematic relationships with others, which would lead them to have a more negative view of others. We discuss this possibility in depth later.

Interpersonal Behaviors

We propose that unmitigated communion individuals’ concern with others’ view of themselves leads them to engage in a variety of “maladaptive” interpersonal behaviors aimed at enhancing one’s self-worth in the eyes of others. We say maladaptive because these behaviors do not lead to satisfying relationships, and they contribute to increased psychological and physical distress. All of the behaviors we discuss are listed in Table 1. Two primary domains of behavior exist: those that reflect overinvolvement with others and those that reflect a neglect of the self. We have shown that both of these behaviors account for the association of unmitigated communion to generalized distress but that overinvolvement with others also accounts for the association of unmitigated communion to situation-specific sources of distress (Fritz & Helgeson, in press, Study 4).

Overinvolvement With Others

Caretaking. A behavior that is common to both unmitigated communion and communion is caretaking. In two longitudinal studies of cardiac patients, both unmitigated communion and communion were related to providing social support over time (Helgeson, 1993; Helgeson & Fritz, 1995). Subsequent regression analyses, however, suggest that unmitigated communion is more strongly related to support provision than is communion. Specifically, in the studies cited earlier, unmitigated communion adds to the prediction of support provision when communion is in the equation, but communion does not add to the prediction of support provision when unmitigated communion is in the equation. In addition, people high in unmitigated communion, compared to communion, seem to carry helping behavior to an extreme. In a longitudinal study on adjustment to heart disease, spouses who scored high on unmitigated communion engaged in more overprotective behavior of the patient over the next 3 months, as reported by both the patient and the spouse, than those who scored low on unmitigated communion (Helgeson, 1993). In studies of college students and adults, unmitigated communion has been associated with self-reports of interpersonal problems that include being overly nurturant, intrusive, and self-sacrificing (Fritz & Helgeson, in press, Study 4). Communion was not associated with these interpersonal problems.

Imbalanced relationships. Another distinction between communion and unmitigated communion lies in the receipt of social support. In two longitudinal studies, one of people with heart disease and one of women with breast cancer, communion predicted support receipt over the next 3 months, but unmitigated communion did not (Helgeson, 1993, 1997). Other research has shown that communion accounts for women’s greater support receipt compared to men’s (Burda, Vaux, & Schill, 1984; Vaux, Burda, & Stewart, 1986).

Thus, people high in communion appear to have more balanced relationships, providing and receiving support, whereas people high in unmitigated communion provide social support without necessarily receiving it. We argue that people characterized by unmitigated communion may prefer this unbalanced state of affairs in which others depend on them but they do not depend on others for support. In one study, unmitigated communion was associated with self-ratings of being uncomfortable receiving support, whereas communion was not (Fritz & Helgeson, in press, Study 2). In order to receive support, one has to be willing to disclose problems. In that same study, unmitigated communion was associated with greater difficulties with self-disclosure, whereas communion was not.

A variety of reasons explain why people who score high on unmitigated communion might prefer imbalanced relationships or feel uncomfortable receiving support. First, this preference could be used as a way to exercise control over relationships, to create a situation in which others depend upon oneself but one does not depend on others. Providing support without receiving support maintains one’s identity as a support provider; receiving support might threaten this identity. Alternatively, the imbalanced relationships may stem from negative self-views. Individuals characterized by un-
mitigated communion might prefer to receive support but do not expect that others will provide support or are afraid that others will not respond to their needs. Thus, they lower their expectations for support to avoid disappointment and rejection. Recall that people high in unmitigated communion report that others do not evaluate them favorably. Such individuals also might feel that they should not burden others with their problems or that doing so would damage their relationships. This last explanation seems to stem from a view that others are more deserving than the self.

**Relationships as a source of self-esteem.** We suggest that relationships have a strong impact on the unmitigated communion individual's self-esteem. We asked 554 college students (334 male, 211 female, 9 unknown) 24 questions about how proud they felt when they provided support to people (e.g., when you help someone solve a problem), how ashamed they felt when they were not at all helpful (e.g., when you have given someone poor advice; when you have been insensitive to someone's feelings), and how ashamed they felt when their help was explicitly or implicitly rejected (e.g., when someone ignores your advice; when someone close to you doesn't seek you out for emotional support). Principal components analysis revealed that these 24 items reflected the three distinct domains described earlier. Both unmitigated communion and communion were associated with feeling good about the self when being helpful \((r = .17, p < .001; r = .19, p < .001)\) and feeling bad about the self when being unhelpful \((r = .44, p < .001; r = .32, p < .001, \text{respectively})\). However, only unmitigated communion was associated with feeling bad about the self when others explicitly or implicitly rejected one's support \((r = .20, p < .001)\). There was no relation to communion \((r = .06, n.s.)\). These two correlations significantly differed, \(t(551) = 3.20, p < .01\) (Glass & Hopkins, 1984). In another study, unmitigated communion—but not communion—was associated with feeling bad about the self when others did not take one's advice (Fritz & Helgeson, in press, Study 2). Thus, a specific relationship difficulty for people high in unmitigated communion may be one in which others do not need them or reject their help.

**Motives for helping.** Although both unmitigated communion and communion are associated with providing support and are correlated with empathy (Fritz & Helgeson, in press, Studies 1 and 3), people high in unmitigated communion and people high in communion have different motives for helping others. We argue that someone who scores high on communion helps out of a genuine concern for others' well-being but that someone who scores high on unmitigated communion helps others to enhance one's own self-worth in the eyes of others. The empathy-altruism hypothesis (Batson, 1995), which states that individuals high in empathy help others in order to benefit the other rather than the self, describes the person high in communion. By contrast, the helping behavior of someone high in unmitigated communion may be better understood by the empathy-specific reward hypothesis or the empathy-specific punishment hypothesis (Batson, 1995). The empathy-specific reward hypothesis states that individuals high in empathy help others in order to obtain rewards for themselves (e.g., others' praise, personal pride), and the empathy-specific punishment hypothesis states that individuals high in empathy help others in order to avoid punishments (e.g., others' disapproval, personal shame).

To test these ideas, we conducted a laboratory study (Helgeson et al., 1997) in which college students were asked to write a story about a time when a friend refused their help and either (a) did not receive help or (b) received help from someone else. We predicted that the person high in communion would be more distressed by a friend not receiving help than by a friend turning to someone else for help because this person is concerned with the friend's needs being met. We predicted that the person high in unmitigated communion, however, would be more distressed by a friend turning to someone else for help than by a friend not receiving help because helping is central to this person's self-concept. The results supported the predictions. In addition, all participants were asked to write a second story in which a friend accepted their help. The unmitigated communion group evidenced greater relief (i.e., improvement in mood) after writing the second story if they first wrote a story about a time when a friend turned to someone else for help. The opposite occurred for the communion group: these people evidenced greater relief after writing the second story if they first wrote a story about a friend not receiving help. In summary, the results suggest that people characterized by unmitigated communion are threatened by a friend seeking help from someone else and are concerned about personally being the help provider, whereas people characterized by communion are concerned about a friend receiving help, regardless of who provides it.

We conclude that helping others may increase the self-esteem of people high in unmitigated communion or communion, but that it may be the primary determinant of how people high in unmitigated communion feel about themselves and not necessarily the primary determinant of how people high in communion feel about themselves.

**Neglect of the Self**

It is not only that the person characterized by unmitigated communion cares for others and provides support
without requiring or wanting support receipt; this individual does so at his or her own expense. Unmitigated communion is associated with a set of interpersonal problems that reflect neglecting one's own needs, such as having difficulty with assertion, being easily exploitable, and inhibiting self-expressiveness to avoid conflict in relationships; communion is not associated with these difficulties (Fritz & Helgeson, in press, Studies 2 and 4). People high in unmitigated communion may fail to assert their own preferences because they believe others would not be responsive. Such persons also may believe asserting preferences could threaten their relationships—relationships that are based on the unmitigated communion person accommodating to others' needs. The self-neglect feature of unmitigated communion is captured in Hare-Mustin and Marecek's (1986) discussion of the negative side of women's relationship orientation: “If relatedness entails the sacrifice of self-interest and precludes women from making claims on others to meet their needs, then it is hard to see how relatedness can promote psychological well-being” (p. 211). We now turn to the implications of self-neglect for the relation of unmitigated communion to well-being.

Implications for Well-Being

Unmitigated communion has implications for psychological and physical well-being, whereas communion does not. First, unmitigated communion is associated with generalized psychological distress that arises in part from the previously described cognitions about the self and, in part, from the two interpersonal behaviors, overinvolvement with others and neglect of the self. Second, unmitigated communion individuals' overinvolvement with others leads to a greater exposure to others' problems that in turn leads to more situation-specific sources of distress. Third, to the extent that a negative view of the self and overinvolvement in others leads to a neglect of the self, physical health problems may result. The differential associations of unmitigated communion and communion to health outcomes are listed in Table 1.

Generalized Distress

Unmitigated communion has been cross-sectionally associated with anxiety and depressive symptoms across an array of healthy populations, including college students (Fritz & Helgeson, in press, Studies 3 and 4), adolescents (Fritz & Helgeson, in press, Study 1), and adults (Fritz & Helgeson, in press, Study 2). In addition, unmitigated communion has predicted depressive symptoms in longitudinal studies of people with chronic illnesses such as adolescents with diabetes (Helgeson & Fritz, 1996), cardiac patients (Helgeson, 1993; Helgeson & Fritz, 1995), and women with breast cancer (Helgeson, 1997). Communion was not associated with psychological distress in any of these studies.

One reason for the association of unmitigated communion to distress is a general lack of self-worth. Low self-esteem is consistently related to depressive symptoms (Taylor & Brown, 1988). We showed in one study that a negative self-view and a reliance on others for esteem are reasons that unmitigated communion is associated with distress (Fritz & Helgeson, in press, Study 1). That study also showed self-esteem mediated the relation of unmitigated communion to depressive symptoms, but that unmitigated communion did not mediate the relation of self-esteem to depressive symptoms. We now turn to a more specific reason for distress.

Situation-Specific Distress

Unmitigated communion is associated with distress in response to specific situations—those that involve exposure to others' problems. Taking care of others and becoming involved in others' problems leads to a greater exposure to stressful events. People high in unmitigated communion become distressed in response to others' adversity and take on others' distress as their own. Individuals who score high in communion or empathy do not. We outline a set of studies that show unmitigated communion is associated with distress because this orientation leads to involvement in others' stressful situations.

Unmitigated communion is associated with self-reports of becoming more distressed when facing situations that involve other people's problems. In a study (Helgeson & Fritz, 1996) of adolescents with Type I diabetes, we assessed stressful life events in four different domains. Unmitigated communion was not related to the number of events in any domain, but was related to being more strongly affected by one domain of events, relationship stressors, whereas communion was not. In addition, relationship stressors accounted for the longitudinal association of unmitigated communion to psychological distress.

Similar evidence was found in a study of 50 female staff members from Carnegie Mellon University (Fritz & Helgeson, in press, Study 2). We asked respondents to review a list of stressful life events (Sarason, Johnson, & Siegel, 1979) twice. First, they indicated whether the event occurred to them and how strongly they were affected by it. Second, they indicated whether the event occurred to someone in their social network and how strongly they were affected by others' events. Unmitigated communion was not associated with the number of stressors that occurred to the self, but was marginally associated with a greater number of stres-
sors that occurred to network members ($r = .27, p = .06$). More important, unmitigated communion was associated with being more strongly affected by events that occurred to others, $r = .44, p < .005$, whereas communion was not, $r = -.11, ns$. These two correlations significantly differed, $t(47) = 3.67, p < .001$. In addition, we asked respondents to think of a stressful life event that happened to someone else and, with that as a reference, to complete the intrusive thoughts subscale from the Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979); that is, respondents indicated the extent to which they were bothered by intrusive thoughts about someone else’s life event. Unmitigated communion was associated with more intrusive thoughts, $r = .35, p < .05$, but communion was not, $r = .04, ns$. These two correlations were marginally different, $t(47) = 1.91, p < .10$.

We conducted two laboratory studies (Fritz & Helgeson, Study 3 and 4, in press) to test the hypothesis that people high in unmitigated communion become more distressed when exposed to another’s stressful life event. In the first study, female college students were exposed to a stranger (confederate) discussing a problem and then contacted two days later to see if they were distressed about the problem. Unmitigated communion was associated with greater distress upon hearing a stranger disclose a problem as well as more intrusive thoughts about that stranger’s problem two days later (Fritz & Helgeson, in press, Study 3). In a replication of that study with friends, unmitigated communion was again associated with intrusive thoughts about a friend’s problem two days later, and the longitudinal association remained significant when baseline distress was statistically controlled (Fritz & Helgeson, in press, Study 4). Communion and empathy were not related to intrusive thoughts over the stranger’s problem or the friend’s problem. Thus, people high in unmitigated communion are susceptible to a situation-specific distress that arises from exposure to other people’s problems. Although this form of distress is distinct from the generalized depressive symptoms that characterize unmitigated communion (Fritz & Helgeson, in press, Study 4), distress upon exposure to others’ problems could contribute to increased generalized distress.

**Physical Health**

Evidence also shows that unmitigated communion has negative implications for physical health, in theory due to the combination of an overinvolvement with others and the failure to attend to one’s own needs. In longitudinal studies, unmitigated communion predicts poor adjustment to a variety of illnesses (heart disease: Helgeson, 1993; diabetes: Helgeson & Fritz, 1996; cancer: Helgeson, 1997). Some suggest that poor recovery from or adjustment to an illness results from the unmitigated communion individual’s neglect of his or her own health. In the study of adolescents with diabetes (Helgeson & Fritz, 1996), unmitigated communion was associated with poor metabolic control (i.e., poor control over blood glucose levels) at 4-month follow-up even when adjusting for baseline metabolic control, presumably because these individuals were not taking care of themselves. Involvement in relationship stressors mediated the longitudinal association of unmitigated communion to poor metabolic control, suggesting that involvement in others’ problems kept people high in unmitigated communion from tending to their own health care. Other research indicates that unmitigated communion is associated with poor health behavior and failure to adhere to physicians’ instructions. In a study of cardiac patients (Helgeson & Fritz, 1995) and an ongoing study of women with breast cancer (Helgeson, 1997), unmitigated communion is associated with self-reports of poor health behavior prior to the onset of the disease. In a study of cardiac patients (Helgeson, 1993), unmitigated communion predicted failure to adhere to physicians’ instructions to decrease household activities over the next 3 months.

Indirect evidence for self-neglect comes from a study of 69 healthy adolescents (Fritz & Helgeson, in press, Study 1). We tested whether visiting the health center was associated with reports of physical symptoms as measured with the somatization subscale from the Brief Symptom Inventory (Derogatis & Spencer, 1982), expecting a positive relation for most people but not for people who scored high on unmitigated communion. If such individuals neglect themselves, they may not respond to symptoms in the way that most people do. There was a positive correlation between physical symptoms and health care visits for those who scored low on unmitigated communion ($r = .47, p < .01$) but not for those who scored high on unmitigated communion ($r = .02, ns$). The difference between the two independent correlations is statistically significant, $z = 2.08, p < .05$. It is also possible that people who score high on unmitigated communion exaggerate their reports of symptoms to attract others’ attention. In that study (Fritz & Helgeson, in press, Study 1), unmitigated communion was marginally related to symptoms, $r = .21, p = .08$.

**Antecedents**

What is the origin of unmitigated communion? Extensive longitudinal data do not exist concerning the origin of unmitigated communion, so admittedly this is a difficult question to address. However, we do have

---

1 To show that unmitigated communion is not related to greater worrying in general, we examined the correlation of unmitigated communion to intrusive thoughts about one’s own problems in another study of 90 undergraduates (Valenti & Helgeson, 1997). We found no relation.
some hints. In our interviews with 50 female staff members from Carnegie Mellon University (Fritz & Helgeson, in press, Study 2), we described unmitigated communion to respondents at the end of the interview and asked where they thought it came from. The most frequent response was the way one was raised (38%). The next most frequent responses were modeling of a family member, usually the mother (19%), lack of self-esteem (16%), and genetics (13%). Our own thoughts on the origins of unmitigated communion are similar. We suspect that unmitigated communion probably results from some combination of genetics (i.e., one may inherit traits such as a communal orientation or low self-esteem with a potential range of expression) and socialization (i.e., the environment shapes those characteristics to the expression of the unmitigated communion trait). In the following we draw on some limited data to speculate about these possibilities.

We suggest that female gender is a risk factor for the development of unmitigated communion. In most studies, but not all, women scored higher than men on unmitigated communion (see Helgeson, 1994, for a review; Helgeson & Fritz, 1996).

Several reasons explain why women may be more likely than men to develop this orientation. First, women are reinforced for involvement in relationships and socialized to regard relationships highly (Miller, 1976). Girls especially may learn that behaving communally is an appropriate way to interact with other people. Women score higher on measures of the communal or expressive orientation than men (Bem, 1974; Spence et al., 1974), and, as previously noted, unmitigated communion is consistently correlated with the communal orientation. Thus, we would not be surprised if women are more vulnerable than men to a kind of communal orientation gone awry.

A second reason that women may be more likely than men to develop unmitigated communion is that women may be more vulnerable than men to negative views of the self. Some evidence shows that women hold more negative views of themselves than men, in that they score lower on self-confidence and underestimate their abilities (see Lenney, 1977, for a review). Women are also more likely than men to make internal attributions for negative events (Boggiano & Barrett, 1991; Hanson & O’Leary, 1985) and to use self-blame as a way of coping with stressful situations (Korabik & Van Kampen, 1995; Vingerhoets & Van Heck, 1990). Thus, when negative events occur, such as others not responding to one’s needs, women may be more likely than men to blame themselves. Self-blame may be adaptive for the person characterized by unmitigated communion because the attribution “there must be something wrong with me” gives one the opportunity to preserve the relationship, whereas the attribution “there must be something wrong with others” does not help to preserve the relationship. Thus, in the context of poor relationships, women may be more likely than men to develop specific behaviors that enhance the quality of their relationships—behaviors that emphasize others at the expense of the self. The combination of an orientation toward others and a negative view of the self may make the idea of involving oneself with others as a way to increase self-esteem a more viable way of behaving for women than men. Thus, low self-esteem may be an antecedent condition as well as a consequence of unmitigated communion.

Regardless of gender or communion, unmitigated communion may be learned in the way that much behavior is learned, through modeling or reinforcement. According to social learning theory, one way people acquire behaviors is by observing others and imitating their behavior (Bandura, 1977). Children growing up in a household with a parent characterized by unmitigated communion may learn that this is an appropriate way to interact with other people and that it is a method to attain self-esteem. In two studies of adolescents (Fritz & Helgeson, in press, Study 1; Helgeson & Fritz, 1996), we found that children’s unmitigated communion scores were correlated with those of their parents ($r = .46, p < .05$, $r = .32, p < .05$, respectively). These data suggest that unmitigated communion behavior may be learned in the home, although it is equally plausible that it is inherited.

Another antecedent of unmitigated communion may be a poor family environment or unsatisfying relationships early in life. We reported earlier in this article that unmitigated communion was associated with insecure adult attachment styles. The extent to which these adult styles are linked to infancy, however, is equivocal (Trelovak, 1997). Thus, no clear data link unmitigated communion to poor relationships early in life. However, we have some evidence that unmitigated communion is associated with poor family relations in adolescence as measured by the cohesiveness and expressiveness subscales of Moos’s Family Environment Scale (Moos, 1974). In a study of teenagers with diabetes (Helgeson & Fritz, 1996), older adolescents who scored high on unmitigated communion appeared to come from households that the parents described as less expressive (i.e., not being conducive to the expression of honest thoughts and feelings), $r = -.58$, $p < .01$, whereas older adolescents who scored high on communion came from households that parents described as more expressive, $r = .47$, $p < .05$. In a study of 69 healthy adolescents (Fritz & Helgeson, in press, Study 1), unmitigated communion was associated with reports of a less cohesive family, $r = -.28$, $p < .05$, and a less expressive family, $r = -.30$, $p < .05$, whereas communion was associated with reports of a more cohesive
family, \( r = .23, p = .06 \), and a more expressive family, \( r = .33, p < .01 \).²

In summary, being a women and being socialized to focus on others may promote the development of communal traits or an orientation toward others. Having poor relationships early in life, and perhaps being female, may predispose one to develop a negative view of the self. Some individuals who have a communal orientation and a negative view of the self will become overly involved with others at the expense of the self; that is, they will develop unmitigated communion, a style of interacting with others that may be modeled or reinforced by the environment.

Final Remarks

We have distinguished unmitigated communion from communion in terms of views of self, views of others, behavior in relationships with others, and psychological and physical well-being. In short, communion is associated with balanced relationships, including support provision and receipt, and satisfying relationships. Unmitigated communion is not associated with satisfying relationships, although unmitigated communion individuals invest heavily in relationships. Unmitigated communion is associated with a negative view of the self. People high in unmitigated communion may focus on others as a way to enhance others’ views of themselves and ultimately their own view of themselves. To this end, such individuals become overly involved in relationships and neglect their own needs. Self-neglect may occur either because people high in unmitigated communion are so involved in helping others that they don’t have time for themselves, because they don’t believe they deserve to attend to their own needs, or because they believe attending to their own needs would jeopardize their relationships. Future research is required to delineate the precise mechanisms by which unmitigated communion is associated with self-neglect. In any case, a cycle is set in motion whereby people high in unmitigated communion attend to others and neglect themselves. We argue that this cycle, as well as this person’s basic negative view of the self, has implications for psychological and physical well-being.

We propose that generalized depressive symptoms are a consequence of unmitigated communion individuals’ low self-regard and reliance on others for self-esteem. However, we also expect that overinvolvement with others and self-neglect lead to increased psychological distress. Overinvolvement with others, in particular, may create situation-specific sources of distress, such as intrusive thoughts about others’ problems, which would then contribute to greater generalized distress. Finally, the neglect of the self, which may stem from a negative self-view or from overinvolvement with others, may have direct effects on physical health.

References

Derogatis, L. R., & Spencer, P. M. (1982) *The Brief Symptom Inventory (BSI): Administration, scoring, and procedures manual*. Baltimore: Johns Hopkins University, School of Medicine
Goldberg, L. R. (1992) The development of markers for the Big-Five factor structure *Psychological Assessment*, 4, 26–42

---

²Because communion and unmitigated communion were positively correlated in both studies \( r = .61, p < .001 \) (Helgeson & Fritz, 1996); \( r = .41, p < .001 \) (Fritz & Helgeson, in press, Study 1), the correlations with communion were examined by controlling for unmitigated communion and the correlations with unmitigated communion were examined by controlling for communion.
UNMITIGATED COMMUNION


Appendix

Revised Unmitigated Communion Scale

Instructions: Using the scale below, place a number in the blank beside each statement that indicates the extent to which you agree or disagree. Think of the people close to you—friends or family—in responding to each statement.

Strongly Slightly Neither Slightly Strongly
Disagree Disagree Agree nor Agree Agree

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I always place the needs of others above my own.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>I never find myself getting overly involved in others’ problems. <em>a</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>For me to be happy, I need others to be happy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>I worry about how other people get along without me when I am not there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>I have no trouble getting to sleep at night when other people are upset. <em>a</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>It is impossible for me to satisfy my own needs when they interfere with the needs of others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>I can’t say no when someone asks me for help.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Even when exhausted, I will always help other people.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>I often worry about others’ problems.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* Items are reverse scored.
Chapter 16
The Social Psychology of Stress, Health, and Coping

Deborah Carr and Debra Umberson

Introduction

That stress affects health is a truism. Laments like “I’m worried sick” convey the conventional wisdom that being “stressed out” harms health. The study of stress and health is one of the richest areas of research in both the social and biomedical sciences, generating hundreds of scholarly studies each year (Thoits, 1995; Wheaton, 1999). The notion that stress makes us sick, anxious, or depressed traces back to the classic book The Stress of Life, in which endocrinologist Hans Selye (1956) wrote that any noxious environmental stimulus would trigger harmful biological consequences. Early social science research similarly argued that any change in one’s social environment, whether positive (e.g., a new baby) or negative (e.g., a death in the family) could overwhelm one’s ability to cope, and increase vulnerability to ill health (Holmes & Rahe, 1967). In recent decades, researchers have moved away from asking whether stress affects health, and have delved more fully into questions of why, how, for whom, for which outcomes, and for which types of stressors does stress affect health. These investigations draw heavily on core concepts of social psychology, and underscore that the extent to which one is exposed to stress, the psychological and structural resources one has to cope with stress, and the impact of stress on health vary widely based on social factors including race, socioeconomic status (SES), gender, age, and psychological attributes including coping style.

In this chapter, we first describe core concepts in the study of stress, coping, and health. Second, we summarize key theoretical perspectives that frame social psychological research on stress and health. Third, we review the methods and measures used, as well as limitations associated with these approaches. Throughout these sections, we draw on examples of empirical studies exploring stressors across multiple life domains, including early life adversity, work, family, and environmental strains, and show their impact on a range of physical and mental health outcomes. We also highlight gender, race, SES, and life course differences regarding the prevalence and nature of stress, coping resources, and stress outcomes. We conclude by suggesting directions for future research on stress, health and coping.
Core Concepts

Stress and Stressors

“Stress” or “stressor” refers to any environmental, social, biological, or psychological demand that requires a person to adjust his or her usual patterns of behavior. Early stress research was conducted on animals, where stress was conceptualized as exposure to noxious environmental stimuli such as extreme temperature (Selye, 1956). Human subjects research, by contrast, typically focuses on social stressors (Holmes & Rahe, 1967; Wheaton, 1999).

Social stressors fall into three major categories: life events, chronic strains, and daily hassles. Life events are acute changes that require adjustments within a relatively short time period, such as job loss. In general, the impact of a stressful life event depends on its magnitude, desirability, expectedness, and timing, where events that are unexpected (e.g., sudden death of spouse) or that happen “off-time” (e.g., being widowed prematurely) are particularly distressing (George, 1999). One subtype, traumatic life events, defined as “extreme threats to a person’s physical or psychological well-being” such as sexual assault or military combat, have especially harmful and lasting effects on health (Thoits, 2010, p. S43). While early perspectives viewed all disruptive life events as distressing (Holmes & Rahe, 1967), contemporary research finds that the impact of an event is contingent on one’s “role history” (Wheaton, 1990), or qualitative aspects of the role one is exiting or entering. Divorce from an abusive spouse, or being fired from an intolerable job may enhance well-being. Conversely, loss of particularly salient and valued roles may especially compromise well-being. A related, but rarely investigated concept is the non-event; recent empirical work shows that not experiencing an event that one had expected, such as marrying or having a baby, can harm one’s mental health (Carlson, 2010).

Chronic strains are persistent and recurring demands that require adaptation over sustained periods, such as a strained marriage, stressful job, or living in a dangerous neighborhood. Chronic strains typically fall into three subcategories: status, role, and ambient strains. Status strains arise out of one’s position in the social structure, such as belonging to an ethnic or racial minority, or living in poverty. Role strains are conflicts or demands related to social roles, such as juggling work and family demands. Ambient strains refer to stressful aspects of the physical environment, such as noise or pollution (Pearlin, 1999). Given their persistent nature, chronic strains are generally found to be more powerful predictors of health than acute events, with the exception of traumatic events (Turner, Wheaton, & Lloyd, 1995).

Daily hassles are minor events and occurrences that require adjustment throughout the day, such as traffic jams, or a spat with a spouse (Lazarus & Folkman, 1984). Historically, most stress research has focused on life events and chronic stressors, although in recent years the collection of daily diary data as a component of population-based surveys has generated interest in daily or “quotidian” strains (Pearlin, 1999). The emotional effects of daily hassles are generally found to dissipate in a day or two (Bolger, DeLongis, Kessler, & Wethington, 1989). Despite the fleeting nature of any one hassle, however, the frequency and type of daily hassles experienced can better explain associated psychological and somatic outcomes than do recent life events or chronic role-related stressors (Bolger et al., 1989). Moreover, daily hassles that recur over long periods of time may become chronic strains and have cumulative effects on health.

Although the three types of stressors often are described as distinctive and discrete experiences, stressors rarely occur in isolation. A life event may create new and multiple chronic strains (e.g., a divorce may create financial strains), and chronic strains may give rise to a stressful life event (e.g., workplace strains may precede involuntary job loss). A stressor in one life domain may carry over to another domain, and a stressor in one person’s life may affect members of his or her social network. Taken together, these patterns are referred to as stress proliferation; this is “a process that places people exposed to a serious adversity at risk for later exposure to additional adversities” (Pearlin, Schieman,
Stress spillover refers to the process where strains in one domain, such as work stress, “spill over” to create stress in another domain, such as one’s family relationships (e.g., Grzywacz, Almeida, & McDonald, 2002). Secondary stressors refer to the strains that emanate following a major life event; for example, a job loss may trigger financial strains (Price, Choi, & Vinokur, 2002).

Scholars have become increasingly interested in the ways that the stressors facing social network members affect one’s own well-being (Kawachi & Berkman, 2001). Network events are stressors facing significant others that spill over into one’s own life; for example, adult children’s divorces may create psychological distress for their aging parents (Greenfield & Marks, 2006). Similarly, stress contagion or stress transfer refers to the process where one person’s reaction to stress affects the health of a significant other, such as when a spouse’s depression following job loss compromises one’s own well-being (Saxbe & Repetti, 2010).

The types of stressors to which one is most susceptible vary widely by one’s social location, reflecting patterns of race, gender, age, and class stratification in the United States. For example, women historically have suffered the “costs of caring” and experienced more stress related to marriage, childrearing, work-family overload, and network events whereas men, on average, have been more vulnerable to financial and job-related stressors; however these differences may converge as men’s and women’s social roles change and converge (Meyer, Schwartz, & Frost, 2008; Thoits, 1995). Ethnic minorities are more likely than whites to experience stressors related to their minority status, including discrimination and interpersonal mistreatment (Meyer et al., 2008), and goal-striving stress (Sellers & Neighbors, 2008). Ethnic minorities, as well as persons of lower SES, are more likely than whites and higher SES persons to experience economic strains, long-term unemployment, poverty, physically dangerous work conditions, and the stressors associated with living in unsafe neighborhoods such as crime victimization (Meyer et al., 2008; Pearlin et al., 2005; Turner & Avison, 2003). Older adults, by contrast, tend to experience stressors related to their own and their spouse’s declining health, caregiving strains, the deaths of spouses and peers, and difficulties negotiating their physical environment, especially following the onset of disability (Zarit & Zarit, 2007).

**Stress Outcomes**

Stress outcomes are the psychological, emotional, or physiological conditions that result from exposure to stress. Early research by Selye (1956) focused primarily on physiological responses to stress, and identified three stages of reaction: alarm, resistance, and exhaustion. Exhaustion, or the depletion of the body’s defenses against stress, was linked to a range of physical health outcomes such as high blood pressure. Most contemporary social psychological studies, by contrast, focus on emotional and psychological adjustments, including depressive symptoms, anxiety, substance use, and self-reported measures of health and illness.

Over the past two decades, a growing number of population-based surveys have obtained biological indicators of health (or “biomarkers”); as such, researchers have become increasingly interested in both physiological indicators of health as outcomes, and physiological responses to stress (e.g., allostatic load) that may contribute to physical and mental health (McEwen, 1998). Allostatic load refers to the physiological consequences of chronic exposure to fluctuating or heightened neural or neuroendocrine response that results from stress exposure.

Most researchers concur that studies should consider multiple rather than single stress outcomes, particularly when comparing stress effects across social groups (Aneshensel, Rutter, & Lachenbruch, 1991). Particular social groups are vulnerable to specific health threats even in the absence of a stressor; thus, focusing on a single outcome may offer potentially misleading findings. For instance, women are more prone to depression and men more likely to use alcohol in the general population,
even when stress exposure is held constant. Thus, studies focusing only on depressive symptoms following divorce may erroneously conclude that divorce affects the well-being of women only; such a study could conceal the fact that men may be more likely to respond to divorce by turning to alcohol, rather than becoming depressed (Horwitz, White, & Howell-White, 1996). Similarly, older adults are believed to have lower levels of stress reactivity, because they have a greater capacity to manage or “regulate” their emotions (Carstensen & Turk-Charles, 1994). As a result, they tend to show less variability in their emotional reactions to stress; for example, older adults tend to evidence less intense and fewer grief symptoms following spousal loss, relative to their younger counterparts (Nolen-Hoeksema & Ahrens, 2001). Studies that focus solely on depressive symptoms or grief, and that neglect a broader range of outcomes including physical health, may erroneously conclude that bereavement is more distressing to young persons than older adults.

Scholars of racial differences in stress outcomes also call for multiple measures, especially given the racial paradox in mental health. Blacks in the United States have higher rates of physical illnesses such as hypertension and diabetes, and higher mortality rates relative to Whites, even after SES is controlled (Williams & Jackson, 2005). However, epidemiologic surveys generally show that Blacks either fare better than or the same as whites in their risk of most psychiatric disorders, including major depression (Kessler et al., 1994; Williams et al., 2007). Researchers disagree regarding the explanations for Blacks’ relatively good mental health, yet many point to methodological issues, including the possibility that standard depressive symptoms scales are culturally biased and may more accurately capture symptoms among whites than blacks (e.g., Breslau et al., 2006; Brown, 2003).

**Coping Resources and Strategies**

The extent to which a stressor affects health outcomes is accounted for, in part, by one’s coping resources and strategies. **Coping** refers to “cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person” (Lazarus & Folkman, 1984, p. 141). **Coping resources** are the personal and social attributes individuals draw upon when dealing with stress (Pearlin & Schooler, 1978). The two main resources identified by social psychologically-oriented stress researchers are social support, and mastery and/or perceived control (Pearlin, 1999; Pearl, Lieberman, Menaghan, & Mullan, 1981). **Social support** refers to the instrumental, emotional, and informational assistance that one draws from others. The number of potential sources of support is less important to one’s well-being than the perception that one can draw on others for support (Wethington & Kessler, 1986). **Mastery** refers to one’s belief that they can control and manage a stressful situation. A high sense of mastery has direct protective effects on health, and also buffers against (or moderates) the harmful effects of stress (Ross & Mirowsky, 1989). However, stress reactions, including psychological distress and depression, may deplete individuals’ usual levels of coping resources when those resources are most needed.

**Coping strategies** are the changes people make to their behaviors, thoughts, or emotions in response to the stressors they encounter (Lazarus & Folkman, 1984). The two main strategies are **problem-focused coping**, where one tries to alter the situation that is causing the stressor (e.g., exiting an unhealthy relationship) or preventing the stressor from recurring, and **emotion-focused coping**, where one alters their reactions to and feelings regarding the stressor, such as finding the humor in the situation (Carver, Scheier, & Weintraub, 1989). Most studies concur that problem-focused tactics are more effective than emotion-focused coping in warding off distress. Problem-focused strategies are associated with lower levels of psychological disorders, whereas emotion-focused strategies are related to higher levels of distress and hopelessness (Billings & Moos, 1981). However, emotion-focused coping may be particularly effective when the stressor cannot be altered, and in the immediate aftermath of the stressor (e.g., Reynolds et al., 2000). The selection and efficacy of a particular coping
strategy is shaped, in part, by one’s coping resources (Lazarus & Folkman) as well as one’s coping style. Coping style refers to one’s general orientation and preferences for addressing problems, such as confronting versus denying (Menaghan, 1983). However, coping styles and strategies alone do not fully determine the health consequences of stress: structural, demographic, and psychosocial factors such as education, social and economic resources, and cognitive flexibility also may moderate whether and how stress affects health (Thoits, 1995, 2010).

Population subgroups vary widely in their access to and reliance on particular coping resources. For example, groups that historically have had less social and economic power tend to have lower levels of perceived control and mastery. Women, ethnic and racial minorities, and persons with lower levels of education tend to exhibit a lower sense of mastery and perceived control, relative to their non-minority counterparts (Turner & Roszell, 1994). However, some historically disadvantaged subgroups have been found to have richer forms of psychosocial support. Women typically report more social and emotional support from friends and children than do men, although men typically receive more support from spouses than do women (Antonucci, 1990).

Evidence is mixed, but some studies conclude that African-Americans have distinctive coping resources that may be particularly effective when dealing with racism and other sources of discrimination; such resources include support from their religious community, protective religious beliefs (Shorter-Gooden, 2004), and high self-esteem (Twenge & Crocker, 2002). A strong sense of racial identity also is a resource that protects against stress, especially racial discrimination. For example, ethnic pride, strong ties to one’s ethnic community, and a sense of commitment to one’s ethnic group protect against distress in the face of discrimination among Filipinos (Mossakowski, 2003) and African-Americans (Sellers & Neighbors, 2008).

Research on subgroup differences in coping strategies reveals clear-cut gender differences, although other subgroup differences, such as race, SES, and age-based differences have not been investigated systematically. Studies consistently show that men and women adopt coping tactics that are consistent with gender-typed expectations regarding emotional display (see Brody & Hall, 2010 for review). Men are more likely than women to use problem-focused coping, control their emotions, accept the stress-inducing problem, not think about the situation, or show emotional inhibition or a “bottling up” of emotions (Lawrence, Ashford, & Dent, 2006; Thoits, 1995). Women, by contrast, tend to seek social support, and use emotion-focused coping tactics such as distracting themselves, releasing their feelings (e.g., crying or talking it out), or turning to prayer (Lawrence et al., 2006; Thoits, 1995). Although social status differences in coping strategies are not well-documented, scholars have argued that ethnic minorities and lower SES persons may rely on strategies that are less efficacious. Pearlin and Schooler (1978, p. 18) observed that “the groups most exposed to hardship are also the least equipped to deal with it.”

Theoretical Perspectives

Several theoretical perspectives, developed by social psychologists, epidemiologists, and sociologists help us to understand the ways that stress affects health, with particular attention to the social structuring of both exposure and responses to stress. The most influential perspectives, including role theory (Biddle, 1979), fundamental cause theory (Phelan, Link, & Tehranifar, 2010), cumulative advantage/disadvantage theory (Dannefer, 2003; Merton, 1968), life course frameworks (e.g., George, 1999), and the stress process model (Pearlin et al., 1981) are undergirded and integrated by the social structure and personality framework (SSP; House, 1977). One of the three “faces” of social psychology (the other two being psychological social psychology and symbolic interactionism), the SSP perspective investigates the processes through which one’s social location, including one’s race, SES, age and gender affects individual outcomes, with particular attention to proximal influences or pathways linking stress to health.
**Role Theory**

*Role theory* holds that most of our everyday activities involve carrying out social roles such as worker or parent. Each social role is accompanied by a set of expectations and norms that guide one’s performance (Biddle, 1979). Role theory was a prominent influence on stress and health research in the 1970s, and was widely invoked as an explanation for why women typically experience more depressive symptoms. Scholars drew attention to the stress created by simultaneously holding multiple roles that taxed one’s coping resources (*role overload*) or that were viewed as in opposition to one another, such as devoted mother and competent worker (*role conflict*). Stress researchers in this era attributed women’s elevated risk of depression to greater exposure to role-related stress, especially juggling work and family (Gove & Tudor, 1973).

Contemporary research counters, however, that juggling multiple roles is not necessarily stressful, nor does it have uniformly detrimental effects on health. First, recent scholarship emphasizes the *salience* (or importance) of the role to the individual. Simon (1992) finds that parenting strains are particularly detrimental for psychological health when the role of parent is highly important and one is highly committed to the role. Second, multiple roles are most deleterious to health when they are involuntary; Ross and Mirowsky (2003) found that for women who wanted to both work for pay and raise children, multiple roles were not particularly distressing. However, full-time mothers who wanted to work for pay, or employed women who wanted to be stay-at-home mothers evidenced elevated psychological distress.

Third, researchers find some evidence for *role enhancement* processes; persons who hold multiple roles may find that difficult stressors in one role are counterbalanced — rather than amplified — by successful experiences in another role (Thoits, 1983). Yet the benefits of role accumulation are not universal, and reflect structural factors including access to high quality and desirable roles. For example, Jackson (1997) found that holding the multiple roles of parent, spouse, and worker provided psychological benefits to whites, but not for Blacks and Puerto Ricans, a pattern that may reflect the relatively poorer quality jobs experienced by racial minorities, as well as work-related stressors such as discrimination or tokenism.

**Fundamental Cause Theory**

Fundamental cause theory (FCT) was initially formulated to explain one of the most persistent findings in social sciences research: the social class gradient in health. SES, whether operationalized as education, income, occupational status, or assets, is inversely associated with nearly all indicators of health, including mortality, self-rated health, disability (i.e., functional limitations), most major diseases and health symptoms, and mental health (e.g., Schnittker & McLeod, 2005). These disparities are stark; for example, persons at the top of the income distribution experience mortality risks roughly half that of those toward the lower end of the income distribution (e.g., Sorlie, Backlund, & Keller, 1995), while life expectancy differs as much as 7 years between higher versus lower income groups (House & Williams, 2000). FCT posits that these gradients persist across a range of health outcomes because SES encompasses a sweeping array of resources, including money, knowledge, power, and beneficial social connections that may affect health regardless of which mechanisms are relevant in a particular context (Phelan et al., 2010). Pearlin and colleagues (2005, p. 207) elaborate that “salient among the circumstances linking status and health is differential exposure to serious stressors” including “the dogged hardships” for which lower SES individuals are at elevated risk.

Empirical studies document that stress partly accounts for the SES gradient in a range of health outcomes. Low SES increases one’s risk of stressful life events ranging from divorce to job loss to early onset of health problems (e.g., Turner et al., 1995), and chronic stressors including poor,
overcrowded and unsanitary living conditions (e.g., Krieger et al., 2002), persistent economic strain (Kahn & Pearlin, 2006), and discrimination (Kessler, Mickelson, & Williams, 1999). SES also is inversely associated with coping resources including supportive social ties (Krause & Borawski-Clark, 1995), the use of adaptive coping strategies (Kristenson, Eriksen, Sluiter, Starke, & Ursin, 2004), and self-esteem (Pearlin et al., 2005). These stressors, in turn, affect health. FCT has been influential because it provides a framework for understanding persistent inequalities in health, and recognizes that there is not a “single bullet” that explains these disparities; rather, multiple pathways including chronic, acute, and ambient stressors play a contributory role.

**Cumulative Disadvantage Theories**

Cumulative disadvantage theories broadly propose that adversity gives rise to subsequent adversity, whereas advantage gives rise to advantage. For example, children who grow up in financially and emotionally secure households have better educational outcomes, which give rise to more stable professional and family lives in adulthood, thus minimizing risk of divorce, job loss, and other health-depleting stressors in adulthood (Gotlib & Wheaton, 1997). As such, an event, experience, or characteristic that has adverse effects in the short-term may take on increasingly vast implications over the life course, leading to a greater bifurcation between the “haves” and “have nots” over time (Dannefer, 2003; Merton, 1968).

Empirical studies provide ample evidence for these cumulative processes, especially regarding the long-term health effects of early life stressors. One longitudinal study of adolescent mothers raised in poverty showed that they were at elevated risk of poverty, family strain, and substance use in adulthood. As more psychosocial risk factors accumulated over the life course, the poorer one’s physical health (measured as functional status) in their 50s and early 60s (Kasper et al., 2008). Similarly, young people who grow up in impoverished households have lower levels of college graduation, steady employment, and earnings as adults (Wagmiller, Lennon, & Kuang, 2008). Those who grow up in families marked by conflict or parental absence also evidence poorer status attainment prospects (Caspi, Wright, Moffitt, & Silva, 1998). As noted earlier, low SES in adulthood and its accompanying strains, are among the most robust predictors of poor physical and emotional health (Phelan et al., 2010). Taken together, these accumulated risk factors help to explain “the ways that …inequalities in health are reproduced” (Thoits, 2010, p. S44).

**Life Course Frameworks**

The life course paradigm, developed by Glen Elder (1995), is a useful framework for studying the origins and impacts of stress (Pearlin et al., 2005). The paradigm has four main themes: (1) human lives are embedded in and shaped by historical context; (2) individuals construct their own life course through their choices and actions, within the constraints of historical and social circumstances; (3) life domains, including work, family, and social background are intertwined; and (4) the developmental impact of a life transition is contingent on when it occurs. Stress studies conducted in the life course tradition have called attention to the importance of timing. The impact of a stressful life transition on psychological health is contingent upon the age or life course stage at which it occurred. For example, one study of the association between birth timing and women’s mental health found that women who gave birth to their first child when they were younger than average (i.e., under age 23) evidenced more depressive symptoms than non-mothers, whereas women who had their first child when they were older than age 23 evidenced fewer symptoms (Mirowsky & Ross, 2002b). Transitions that happen
“earlier” than is typical are viewed as particularly stressful because one may lack the peer support, maturity, life experience, or education essential to managing the stressor.

Similarly, the health impact of chronic and acute strains may vary based on one’s birth cohort, given that the meaning of a particular experience, such as work-family strain, may vary based on the sociohistorical context in which one was raised. For example, juggling simultaneous work-family demands (versus being a full-time parent) takes a more deleterious emotional toll on women belonging to birth cohorts who were raised to prioritize a “traditional” gendered division of labor, where women devoted their energies to raising their children rather than pursuing careers (Carr, 2002).

The life course paradigm has also been influential in underscoring the importance of childhood events and conditions for later-life health and well-being. Early studies of stress were founded on the assumption that only recent or current stressors would affect physical and mental health (Wheaton, 1999). However, empirical work now shows persuasively that childhood and adolescent experiences including parental death (Slavich, Monroe, & Gotlib, 2011), parental divorce (Amato, 2000), child abuse victimization (Slopen et al., 2010), poverty (Duncan, Ziol-Guest, & Kalil, 2010), and living in an unsafe neighborhood (Varatian & Houser, 2010) have deleterious implications for adult health.

As discussed above, one pathway linking early adversity to adult health is cumulative disadvantage processes, where early adversity exposes one to a range of subsequent adversities and health-depleting social stressors. Yet recent studies have found that early life adversities also may trigger physiological responses that impede health in both the short- and longer-term (see Taylor, 2010 for review). Exposure to early life stress has been found to affect the body’s two major stress systems, the sympathetic nervous system and the hypothalamic pituitary-adrenal (HPA) axis. Sympathetic arousal leads to the secretion of hormones such as catecholamines, which may trigger stress-related changes in blood pressure and heart rate. Dysregulated HPA functioning also is linked with long-term consequences for health and functioning, including diabetes and asthma risk. Early life adversity, such as growing up in a harsh family, also has been found to affect neural functioning; each of these physiological reactions puts one at an elevated risk of multiple health problems over the life course.

The Stress Process Model

The stress process model is the most influential framework for contemporary scholars exploring stress, coping, and health (Pearlin et al., 1981), in part because it incorporates key themes of FCT, cumulative disadvantage, and life course frameworks. The model holds that most stressors are rooted in roles that link individuals to social structures and that are allocated, in part, on the basis of characteristics like age, race, and gender. Exposure to stress is not randomly distributed throughout the population, but is highly structured and reflects patterns of inequality. Additionally, the impact of a stressor on health is not universal in magnitude, and varies widely based on one’s other risk factors and resources, such as social support, self-esteem and mastery. Many of the key themes and concepts of stress research described in this chapter are derived from Pearlin’s original model (1981) and later elaborations (Pearlin, 1999). The model holds that stressors may take multiple forms (events, chronic, quotidian) and consequences may manifest in multiple ways, thus affecting a range of physical, emotional, and interpersonal outcomes.

Research Methodologies

The dominant method for studying stress, health and coping is the analysis of data from large-scale surveys, yet in recent years these data have been supplemented with biomarker, daily diary, observational, and qualitative data. Heightened interest in the use of “mixed” or “blended” methods has enabled to researchers to not only document patterns, but also to explore the processes through which stress affects health, and the symbolic meanings of particular stressors.
Quantitative Research Approaches

Data Sources

Help-Seeking Samples

The data and methods available for sociological studies of stress and health have undergone important transformations over the past five decades. Early studies drew subjects from help-seeking populations, such as those seeking psychiatric treatment for grief following spousal loss (Parkes, 1965) or parents receiving counseling for themselves or their children following divorce (Wallerstein & Kelly, 1980). These studies may overstate the negative impact of stress because persons with the most difficult readjustments are over-represented.

Cross-Sectional Studies

The next wave of studies relied on cross-sectional survey data, which provide a single point-in-time “snapshot” of a population. Such samples allow researchers to compare the health of persons who have versus have not experienced a particular stressor. However, cross-sectional data have important limitations. First, researchers cannot disentangle causation and selection (e.g., Goldman, 1994). Social causation posits that a stressor, such as divorce, causes negative outcomes such as depression. Social selection, by contrast, proposes that an observed difference between two subgroups may not be due to the stressor per se, but to differences that existed prior to the event or onset. For example, economic strains, substance use, or depression may give rise to a marital dissolution; thus, higher levels of depression among divorced persons may reflect difficulties they would have faced even in the absence of the dissolution. Second, cross-sectional data do not allow researchers to easily identify the direction of causation, or whether causation is mutual (i.e. marital strain increases depression, and vice-versa). Third, cross-sectional data do not allow researchers to document the impact of stress on health trajectories over time. Studies that focus on a single time point tend to reveal only the short-term consequences of a stressor, and fail to detect health consequences that are lagged and emerge long after the stressor has ended. This is particularly important for studying physical health consequences, as many chronic diseases have long latency periods. For example, Zhang and Hayward (2006) found that divorce increased heart attack risk in women decades after the dissolution occurred.

Longitudinal Studies

Longitudinal studies are superior in revealing causal influences, because they can better pinpoint the temporal ordering of events and experiences. Multiple data points are particularly important when exploring the consequences of stressful life events; most discrete events take time to come to fruition and often occur after a period of chronic stress (Avison & Turner, 1988). For example, job loss may occur at the end of a long period of uncertainty about one’s employer (Burgard, Brand, & House, 2009).

Longitudinal studies have limitations, however, including a high financial cost, and selective attrition (i.e., loss of particular subjects over time). The high drop-out and death rates of unhealthy persons may lead researchers to underestimate the potentially harmful health consequences of a stressor. Stress researchers should identify both the sources and possible consequences of sample attrition. More sophisticated strategies, such as weighting adjustments, imputation (Little & Schenker, 1995), and the estimation of two-stage selection models (Heckman & Singer, 1984) are effective ways to begin to address the issue of selective attrition.
Sample Surveys

Large sample surveys enable researchers to explore subgroup differences in the extent to which stress affects health, paying attention to race, gender, and SES differences in the magnitude, direction, and potential explanations for these effects (e.g., George & Lynch, 2003). Further, most sample surveys are designed to study a wide array of general topics, and include a range of health and stress measures, as well as sociodemographic and economic characteristics that might account for a spurious association between stress and health.

However, the general nature of most surveys can also be a limitation when studying specific stressors. Some stressors are so rare that even a survey with 10,000 respondents may not be sufficient for adequately powered statistical analyses. For instance, suicide rates in the United States are low, so researchers hoping to identify the effects of a spouse’s suicide on survivor health may have too few cases to run meaningful analyses. General surveys also may not capture secondary stressors or contextual factors related to specific stressors. For example, surveys used to study the health implications of divorce tend not to obtain information on stressors or contextual factors specific to the transition, such as negotiating complex residential and financial relationships with ex-spouse, new spouse, and stepchildren (Carr & Springer, 2010).

Analytic Approaches

Researchers using survey data to study stress and health typically use multivariate statistical methods which allow them to build sequential, nested models that reveal the causal pathways or mediators linking stress to health. Moderation analyses, or interaction terms, allow researchers to identify the joint effects of two co-occurring stressors, or to compare the effect of a particular stressor on two different subgroups, such as men versus women.

Given the threats to causal inference described above, researchers also are using sophisticated quantitative methods to help tease out causal ordering. Structural equation models (or path analysis) allow researchers to identify mutually influential pathways. Propensity score matching, by contrast, allows researchers to adjust for selection bias. Massoglia (2008) used this method to show that the stress of imprisonment increased men’s risk of infectious disease, and that these effects were not plausibly due to the myriad other stressors that preceded incarceration. Researchers interested in documenting the ways that stressors affect health trajectories over time may use latent growth curve models, which allow evaluations of the duration, course, and patterning of health symptoms. For example, Haas (2008) found that poor childhood health and disadvantaged social origins were associated with both more functional limitations among older adults at a baseline interview and accelerated increase in limitations over time.

Measurement Issues

Insights from social psychological theory and research point to limitations of past research and have driven recent advances in measurement for stress research.

Stress and Stressors

Stressful Life Events

Early measures of stressful life events relied primarily on checklists of stressful events that were weighted in terms of severity (Holmes & Rahe, 1967). Some stress researchers still rely on checklists
of events experienced in a particular time frame, and use simple counts of recent events, as they are significant predictors of disease and distress (Cohen, Janicki-Deverts, & Miller, 2007). However, event counts do not allow researchers to identify the precise pathways linking a distinctive stressor to health outcomes. The highest quality contemporary research on stressful life events relies on population-based studies that include detailed measures of physical and mental health, as well as specific stressful life events (recent life events as well as life-time traumatic events) and chronic strains. Measures of acute events often are derived from survey questions designed initially for other purposes. For instance, widowhood and divorce are obtained in marital histories, births are obtained via child rosters or fertility histories, school dropout in educational histories, and job losses captured in occupational history measures. When using such indicators of events, however, researchers must take great care to identify the number of months or years elapsed between the time at which the event occurred and the time when the health outcome is measured.

Chronic Strains

Chronic strains, like stressful life events, can be measured with inventories designed expressly to capture strain such as Wheaton’s (1994) 51-item inventory of chronic strain, or from survey items designed for other purposes that capture persistent stressful experiences. Among the chronic strains measured on surveys that have been linked to health outcomes include time pressures on the job (Roxburgh, 2004), perceived job insecurity (Burgard et al., 2009), marital strain (Umberson, Williams, Powers, Liu, & Needham, 2006), intimate partner violence (Campbell, 2002), parent–child relationship strain (Greenfield & Marks, 2006), work-family conflict (Frone, 2000), caregiving (Lee, Colditz, Berkman, & Kawachi, 2003), institutional and interpersonal discrimination (Kessler et al., 1999), and financial strain (Kahn & Pearlin, 2006). Researchers have become increasingly interested in evaluating the long-term health impact of chronic early-life stressors such as child abuse (Springer, Sheridan, Kuo, & Carnes, 2007); living with a depressed parent (Repetti, Taylor, & Seeman, 2002), childhood poverty (Montez & Hayward, 2011), and parental marital strain (Amato, 2000).

Most measures of chronic strain are self-reported by study participants. However, for environmental and contextual stressors, researchers may link individual-level survey data with area- or neighborhood-level data capturing strains such as poverty (Krieger et al., 2002), crime rates (Sampson, Morenoff, & Gannon-Rowley, 2002), or proximity to industrial activity (Downey & Willigen, 2005). These area-level indicators are typically captured at the Census tract or neighborhood level, and are appended to an individual’s survey record. For example, the proportion of persons in one’s Census tract living beneath the poverty line is a powerful predictor of one’s physical health (Krieger et al., 2002).

Daily Hassles

Advances in daily diary methods have contributed to a recent flourishing of research on daily hassles/strains. Contemporary methods ask respondents to report on hassles and uplifts at the end of each day (or even multiple time points each day) for a number of consecutive days or weeks. Diary methods typically have participants respond over the telephone, with personal digital assistants, or on Internet sites (Almeida, 2005). These methods are particularly useful for examining daily fluctuation in stress in relation to psychological and physical symptoms.

Stress Outcomes

The main outcomes we focus on in this chapter are mental and physical health, and health behaviors. As noted earlier, stress researchers generally agree that multiple outcome studies are superior to single
outcome studies (Aneshensel et al., 1991). This focus on multiple outcomes allows researchers to identify the distinctive health consequences of stress for particular subgroups, and pinpoint those outcomes that are not affected by a purported stressor. Inconsistencies across outcomes will ultimately lead to more refined theories (Thoits, 1995).

We discuss mental and physical health as separate sets of outcomes in this chapter, as most researchers tend to focus on one broad set of outcomes only, based on their expertise and training. However, the two are often mutually influential. Depression is an important pathway through which stress affects physical health; Finch and colleagues (Finch, Hummer, Kol, & Vega, 2001) found that depression is the primary mechanism through which race-based discrimination affects the symptom counts and self-rated health of Mexican Americans. Conversely, physical health problems are a consistent predictor of psychological health, reflecting pathways including physiological factors, strains of managing chronic illness, and the detrimental impact of chronic illness on the quality of life. For example, Pudrovska (2010) finds that cancer increases depression risk, in part, because it poses a challenge to one’s identity and sense of control, especially for men.

Mental Health

Given that most empirical studies of health and stress draw on survey data, the mental health outcomes considered are those that can be easily and accurately measured using self- or telephone-administered (rather than clinician-administered) instruments. The most commonly studied outcome is depressive symptoms, typically measured with the Center for Epidemiological Studies Depression (CES-D) scale (Radloff, 1977). Major depressive disorder (MDD) also is commonly studied, using measures such as the Composite International Diagnostic Interview (CIDI, Robins et al., 1989). Other psychological outcomes include positive and negative affect (e.g., PANAS, Watson, Clark, & Tellegen, 1988), and anxiety (e.g., Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). Studies of bereavement also focus on symptoms of grief and loss-related distress.

Most measures evaluate symptoms that one has experienced in the past week or month, and require one to make an aggregated assessment of their feelings during that time period. However, scholars of well-being have recently called for heightened attention to an alternative measure: experienced well-being, or the momentary reports of how one is feeling. Recent studies suggest that a more immediate and momentary measure of well-being that assesses one’s mood or level of physical pain during a randomly selected time period in the day prior to interview provides a more accurate snapshot of immediate stress response, and one that is not subject to recall bias (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Steptoe & Wardle, 2011).

A persistent debate in stress research is whether outcomes are best treated as dichotomies, indicating the presence or absence of a clinical condition, such as MDD, or as continuous measures such as number of depressive symptoms experienced in the past week (Horwitz, 2002). Dichotomous outcomes are consistent with clinical practice, where treatment and reimbursement are based on formal diagnosis (Horwitz, 2002; Kessler, 2002). However, most social psychologists emphasize that stress outcomes should be conceptualized and measured more broadly, as both continua and discrete categories. Studies focusing only on dichotomous outcomes, may underestimate the health consequences of a stressor; distressed individuals who barely fail to meet the criteria for the diagnosis are disregarded (Mirowsky & Ross, 2002a).

A focus on psychological disorder also precludes attention to positive mental health. Ryff and Singer (2001) observe that persons who score very low on indicators of positive psychological adjustment, such as positive mood, may find themselves at an elevated risk of MDD if confronted with additional stressors. Although, on average, stress undermines mental and physical health, emerging research on “post-traumatic growth” also suggests that, for certain individuals in certain social contexts, stressors may give rise to positive mental health especially personal growth, as individuals
recognize that they can survive and thrive in the face of difficulties (Tedeschi & Calhoun, 2004). However, these effects typically are not evidenced until significant time has elapsed since the stressful period. For example, Carr (2004) found that widows who had been most dependent on their spouses during marriage evidenced the greatest increases in self-esteem and personal growth postloss.

Physical Health

Most studies of stress and physical health focus on general outcomes such as self-rated health, number of current illnesses or health symptoms, and all-cause mortality. General outcomes are used because they are widely available in population-based studies and are appropriate for studying all age groups and both genders. Specific health conditions also may not have a sufficiently high prevalence rate in a particular survey to allow multivariate analyses. Among the most commonly used specific health outcomes are heart disease, high blood pressure, and other health outcomes that are plausibly linked to stress and are sufficiently high prevalence to allow multivariate analyses.

A promising research avenue is the investigation of physiological pathways through which stress “gets under our skin,” with particular attention to cardiovascular, endocrine, immune, metabolic, and sympathetic nervous systems (Ryff & Singer, 2001, p. 214). Laboratory-based studies of marital strain consistently show that stressful conflict can impair immune response, slow wound healing, heighten susceptibility to infectious agents, increase cardiovascular reactivity, and ultimately compromise physical health (Robles & Kiecolt-Glaser, 2003). Although this research historically was confined to the laboratory and conducted by psychologically-oriented social psychologists, the collection of biomarker data in large-scale surveys now enables SSP-oriented researchers to explore physiological pathways. Recent work also shows how this stress process begins in childhood and has cumulative physiological effects over the life course (Repetti, Robles, & Reynolds, 2011).

Several large representative sample surveys of adolescents (e.g., Add Health) and adults (e.g., Midlife in the United States [MIDUS]) have supplemented self-reported health data with biomarker measures. Studies based on the MIDUS show that exposure to stressors including child abuse (Slopen et al., 2010), discrimination (Friedman, Williams, Singer, & Ryff, 2009), and marital strain (Whisman & Sbarra, 2012) elevate one’s inflammation levels at midlife; inflammation is a well-documented correlate of cardiovascular disease and cancer. Stress researchers also have expanded their foci to include cellular aging; both current stressors as well as those dating back to prenatal conditions may speed up cellular aging, operationalized as shortened telomere length. More rapid cellular aging has been detected among young-adult children of mothers who were exposed to psychological trauma during pregnancy (Entringer et al., 2011) and persons who anticipate stressful events in the near future (O’Donovan et al., 2012).

Health Behaviors

Health behaviors, including smoking, drinking, exercise, and sleep are important stress outcomes in their own right, and also are a critical pathway linking stress to physical and mental health outcomes. Health behaviors vary widely by gender, race, and SES, and as such, are an important mechanism in understanding subgroup differences in health (see Schoenborn & Adams, 2010 for review). In general, persons with lower levels of education and income are more likely than their higher SES counterparts to smoke and engage in problematic drinking, and are less likely to maintain healthy diets, exercise regularly, and maintain a healthy body weight. Women are less likely than men to smoke and drink have, although these gender gaps have narrowed among recent cohorts. Women typically are at greater risk for overweight and obesity compared to men, however. African American adults have smoking rates similar to whites, yet engage in less frequent exercise and are at greater risk
of obesity, especially African American women (Schoenborn & Adams). Race differences in alcohol use are complex, however; a higher percentage of African American and Latino adults abstain from alcohol use but also a higher percentage among those who drink are heavy drinkers, relative to whites (Galvan & Caetano, 2003).

Recent work merges stress and life course perspectives to demonstrate how stress influences a range of health behaviors, depending on one’s life course stage. For example, stress contributes to weight gain for adults under age 55 but to weight loss for older adults (Umberson, Liu, & Reczek, 2008). Stress exposure has been associated with unhealthy diets (Ng & Jeffery, 2003), smoking (Steptoe, Wardle, Pollard, Canaan, & Davies, 1996), alcohol consumption (Steptoe et al., 1996), physical inactivity (Ng & Jeffery), and poor sleep quality (Burgard & Ailshire, 2009). Some health behaviors reflect coping strategies; overeating, smoking, and drinking may alleviate psychological/physiological arousal and regulate mood state, at least temporarily (Kassel, Stroud, & Paronis, 2003). Health behaviors, in turn, are associated with negative health outcomes, including heart disease, depression and cancer. Recent work also emphasizes that stress affects health behavior of children and adolescents, contributing to cumulative disadvantage in health over the life course (Repetti et al., 2011).

Coping Resources

Social psychological studies of stress and health have been strongly influenced by the stress process model (Pearlin et al., 1981), and focus on two main resources: social support, and a sense of mastery or perceived control over one’s environment. Analytically, coping resources usually are treated as either a mediator of the impact of stress on health, or a moderator that “buffers” against its health-depleting effects (Pearlin et al.). However, researchers also acknowledge that coping may affect stress; coping resources may prevent a stressor from occurring in the first place or from spiraling out into secondary stressors (Wheaton, 1999).

Diverse measures are used to capture social support and mastery. Social support refers to the functions performed for an individual by significant others, including family, friends, and colleagues. The types of support provided may be instrumental (e.g., financial), emotional (e.g., listening to one’s problems), or informational (e.g., providing advice). Early studies focused on structural aspects of potential support, such as the number of persons in one’s social network. Most research concurs, however, that subjective aspects of social support are more protective than simple counts of significant others (Wethington & Kessler, 1986), although the two are highly correlated (House & Kahn, 1985). A very simple measure, asking whether one has a person with whom one can share their private thoughts has proven to be a powerful predictor of health (Cohen & Wills, 1985).

Multi-item scales capturing positive and negative relationship characteristics also are widely used (Rook, 1998). Positive aspects include feeling loved and understood, and negative aspects include criticism and conflict (e.g., Multidimensional Scale of Perceived Social Support, MSPSS, Zimet, Dahlem, Zimet, & Farley, 1988). “Ambivalence,” which refers to having both positive and negative sentiments toward a single person or relationship also carries health implications (Pillemer & Suitor, 2008).

An equally important coping resource is a sense of mastery, or control over one’s environment. An early measure, locus of control, captured one’s attributions for the events in his or her life (Rotter, 1966). An individual could attribute a personal experience to his/her own actions or characteristics (internal) or to situational factors (external). More recent measures capture different types of control beliefs, such as the personal ability to control the events and experiences of one’s life (personal control or self-efficacy). Commonly used measures of self-efficacy include Ryff’s (1989) environmental mastery and Bandura’s (1977) self-efficacy.
Qualitative Research Approaches

Studies based on population-level data have clearly established that stress has measurable effects on individuals in the general population, identified the psychosocial factors that protect and exacerbate the effects of stress on health, and revealed group differences (e.g., SES, gender, and race) in exposure to stress. However, quantitative methods are limited in their ability to reveal the underlying psychosocial processes and symbolic meanings linking stress to health. Qualitative methods are designed to directly study these processes and meanings. Qualitative methods involve in-depth intensive study of specific cases, in contrast to quantitative methods which require a large number of cases to generate links between variables. Qualitative methods are particularly important when social scientists are breaking new ground and when studying difficult-to-reach populations (Ragin, Nagel, & White, 2004).

The most commonly used qualitative approach in stress research is in-depth interview methods. Researchers identify issues of process and meaning to be addressed and recruit individuals for in-depth analysis. Qualitative methods then provide data to analyze the meanings and processes through which specific stressors shape mental and physical health as well as health behavior. For example, in-depth interviews and observations have been used to identify the specific nature and consequences of workplace stressors (Peterson et al., 2010).

Blended-Method Approaches

While quantitative and qualitative studies provide independent contributions to the study of stress and health, blended methods are particularly fruitful for investigating stress/health linkages. Quantitative and qualitative data have distinctive strengths than can build on one another (and partly address the limitations of each strategy) to provide a more complete understanding of social psychological processes (Axinn & Pearce, 2006). For example, Umberson (2003) presents quantitative results from a national longitudinal survey to show that the death of a parent in adulthood is associated with a significant decline in health over time. Qualitative data from in-depth interviews with recently bereaved adult children then reveal psychosocial processes that occur following the death and how those processes influence health habits that affect health outcomes. The qualitative analysis shows that adult children strongly identify with their deceased parent; the death of a parent leads adult children to become more aware of their own mortality and improve their health habits in an effort to live longer and in better health.

Used in tandem, qualitative methods can reveal psychosocial processes that account for the patterns revealed in population level data and, in turn, quantitative approaches can be used to document these processes at the population level. Qualitative findings also may generate new questions that can be addressed at the population level and suggest specific measures that should be included in future quantitative data collections (Ragin et al., 2004). Ideally, blended methods investigations should move back and forth between quantitative and qualitative strategies so that the result is a richly nuanced understanding of how structure and meaning coalesce to explain the impact of stress on individuals (Axinn & Pearce, 2006; Pearlin, 1999). While we have emphasized the blending of population level analyses with in-depth interview analyses, other strategies also are fruitful. For example, qualitative data might reveal psychosocial processes linking relationship stress to health habits. Daily diary methods could then provide a quantitative assessment to reveal how those psychosocial processes unfold in daily interactions with others.
Future Directions

We propose four avenues that we believe are fruitful areas for future research on the social psychology of stress and health: heterogeneity in the impact of minority stress; consideration of dyad, family, and network-level appraisals of stress; use of multiple outcomes and measurement modes; and attention to genetic moderators of the stress-health relationship.

Heterogeneity and Minority Stress

Researchers of stress and health have made tremendous strides in identifying the distinctive stressors, coping resources, and stress outcomes experienced by members of minority groups, broadly defined. We encourage the continued exploration of sources of heterogeneity even within specific subgroups. The concept of “minority stress” holds that members of minority groups, whether due to race, ethnicity, or sexual orientation face stigma, discrimination, and prejudice, and these processes create a hostile and stressful environment that can elevate one’s risk of mental and physical health problems (Williams, Yu, Jackson, & Anderson, 1997). However, some studies suggest that some members of minority groups are resilient in the face of stress. For instance, lower rates of depression and other mental health symptoms have been observed among African Americans relative to whites. This advantage has been attributed to a range of explanations including measurement issues, such as Blacks’ tendency to respond to somatic versus affective items in standard depression scales, and the protective resources enjoyed by African Americans including religious coping and reliance on extensive support networks including friends and “fictive kin” (Williams et al., 1997). However, recent work also suggests that Blacks may be more likely to respond to stress with risky health behaviors that reduce psychological distress in the short-run even while contributing to poorer health in the long-run (Jackson, Knight, & Rafferty, 2010). Future studies should delve more fully into the distinctive risk factors and psychosocial resources of populations historically subject to “minority stress”, including immigrants, gays, lesbians and bisexuals, and obese individuals who may be subject to stigma and discrimination. Future studies should use longitudinal data to document the short- and longer-term consequences of minority stress, and should focus on multiple outcomes to pinpoint the precise ways that minority stress affects the well-being of distinctive subgroups.

Moving Beyond the Individual

Many of the most commonly studied stressors, such as family transitions, relationship strains, poverty, and environmental strains, occur at the dyadic, family, or social network level. However, one of the most ironic limitations of studies on “social” stressors and health is that most focus on one individual within the larger social network. This limitation is due, in part, to traditional modes of data collection where one person reports on his or her own union and parental statuses, relationship quality, and self-rated health as well as the health of one’s spouse or a randomly selected child. Although studies based on such data are immensely valuable in documenting associations and causal pathways, they fail to capture the complexities of social life – including the possibility that two romantic partners, siblings, co-parents, coworkers, or neighbors experience their social context (and its health consequences) in starkly different ways.

Dyadic data analysis allows researchers to use data from multiple reporters, such as husbands’ and wives’ reports of marital strain, to estimate how much each person’s outcome is associated with both
own and partner characteristics. This approach enables researchers to explore how both spouses’ reports of marital conflict are associated with each spouse’s health outcomes (Heffner et al., 2006; Sandberg, Harper, Miller, Robila, & Davey, 2009), for example. We suspect that these pathbreaking studies and methods will set the stage for more nuanced studies of social stressors, their impact, and potential proliferation.

**Reconciling Multiple Outcomes and Methods**

Considering multiple stress outcomes is an essential step to understanding the health impacts of stress (e.g., Aneshensel et al., 1991); however, we urge researchers to also consider diverse modes of measuring these outcomes, including biomarker and self-reported measures of physical health, as well as general/aggregated versus momentary measures of psychological health. We also encourage researchers to expand their analyses of general health measures (e.g., all-cause mortality, self-rated health) to include specific and high-prevalence outcomes, such as specific health behaviors, risk of heart disease, high blood pressure other relatively common conditions that can be studied at the population level.

Current research on cardiovascular disease provides an exemplar of how knowledge is successfully accumulated across measures and methods. Survey data reveal that divorced persons are more likely than their married peers to die from heart attacks and have a poorer likelihood of recovery after receiving a diagnosis of cardiovascular disease (e.g., Idler, Boulifard, & Contrada, 2012). Laboratory and biomarker studies show that persistent high quality emotional and instrumental support both reduce risk of a coronary event and facilitate recovery (Robles & Kiecolt-Glaser, 2003). Small-scale qualitative studies have identified modifiable factors such as spouse and patient strains and fears that are associated with poor recovery from a coronary event (e.g., Santavirta, Kettunen, & Solovieva, 2001). This kind of cumulative knowledge building, through the use of multiple data sources and methods sets an example for future studies of stress and its influence on the etiology, onset and progression of mental and physical health conditions.

**Gene-Environment Influences**

Researchers have long attempted to understand the relative contributions of genetic versus social influences on health. In the last decade, however, scientific knowledge and available data have become sufficiently sophisticated to accurately identify specific gene/environment interactions that affect health. One line of research builds on early sibling studies, but uses new data sources (e.g., survey data on adopted, biological, and twin siblings) and modeling techniques (fixed- and random-effects models) to assess the distinct contributions of genetic factors and shared social stressors on health outcomes (Pudrovská, 2008).

A highly promising development is the identification of specific genetic polymorphisms (i.e., genetic variations that produce different outcomes within the same species) that affect health risks both directly and in conjunction with stress. For example, Caspi and colleagues (2003) found that a specific polymorphism in the promoter region of the serotonin transporter (5-HTT) gene moderated the influence of stressful events on young adults’ depression and suicidality. These provocative findings suggest that a genetic predisposition may heighten or suppress the health impact of a stressor, and conversely, a stressor may enhance or suppress the effect of a genetic propensity on one’s health. Future research may reveal both those individuals at greatest genetic risk of health problems, and the stress processes that exacerbate (or protect against) these risks. Despite the potential of behavioral
genetics research to uncover pathways linking stress and health, we caution researchers to carefully assess the policy implications of this work. Studies revealing that a particular subgroup has a genetic predisposition for a particular mental or physical health condition raises the potential for stigmatization of individuals who belong to that group (Duffy, 2000). In the worst case scenario, genetic explanations might be used to support racial, gender, and other stereotypes regarding the superiority or inferiority of specific subpopulations.

Taken together, research in these four realms will help to clarify why, how, for whom, for which outcomes, and for which types of stressors stress affects health. Ultimately, social psychological research on stress and health has high potential to identify potentially modifiable factors, and to generate policies and practices to minimize persistent social inequalities in health.

References


Horwitz, A. V. (2002). Outcomes in the sociology of mental health and illness: Where have we been and where are we going. *Journal of Health and Social Behavior, 43,* 143–151.


Age, birth cohort, and period effects in psychiatric disorders in the USA

Katherine M. Keyes and Charley Liu

5.1 Introduction

Risk for psychiatric disorders varies across the life course, with incidence most likely to occur during specific critical periods of development. As the life course perspective increases in utility to understand disease aetiology, analysis of period and cohort effects in historical data over time has become imperative for understanding the role of biological and environmental factors in disease distribution at the population level. Temporal variation in rates of disease may reflect changes in the nature and magnitude of the effect of aetiologically important exposures, and identification of particular cohorts with higher than expected rates of psychiatric disorders is vital for public health planning and intervention efforts.

In the present chapter we review three disorders with sufficient research bases in age, period, and cohort effects such that conclusions can be generated: major depression, alcohol use disorders, and autism spectrum disorders. We focus our review on studies conducted in the USA, as trends in psychiatric disorders and related mortality vary across countries and cultures; thus a cohesive framework for investigation of age-period-cohort effects is best suited in defined geographic areas. We begin by defining and describing age, period, and cohort effects and statistical considerations in estimating and interpreting models.

5.2 What are age, period, and cohort effects?

Age effects can be conceptualized at both the individual and the ecological level. Individual-level age effects are developmentally linked pathways to health outcomes; for example, individuals are most likely to initiate cigarette use between the ages of 13 and 15 years, and alcohol use between the ages of 15 and 17 years. Ecological-level age effects refer to the age structure of the population (i.e., the distribution of older and younger individuals over time). For example, the prevalence of Alzheimer's disease is expected to increase as the proportion of individuals aged ≥80 years in the USA increases. Period effects describe changes in the prevalence of health outcomes associated with certain years, and these affect all age groups simultaneously. Common sources of period effects are changes in law or policy, introduction of an environmental pollutant in the population, or other factors that could affect all age groups simultaneously. Cohort effects describe changes in the prevalence of an outcome associated with certain age groups in certain years. That is, cohorts share exposure to certain environmental conditions at critical developmental ages, they compete for positions in labour and educational markets, and they experience historical events such as war and recession at the same age. Cohort effects in a health outcome indicate that the shared experience of early childhood, coming of age, and growing older with others of the same age can shape health.
5.3 Design and analysis considerations in age–period–cohort modelling

5.3.1 Study design: cross-sectional

Evaluation of birth cohort effects began as a life table method dating as far back as Farr, and graphical approaches were further developed throughout the first half of the 20th century. Whereas many studies use repeated cross-sections of data taken over time for age–period–cohort research, such data are often unavailable especially for psychiatric disorders, as criteria for diagnoses have shifted over time. Thus, single cross-sectional studies comparing retrospectively reported lifetime prevalence of disorder across individuals in different cohort groups are commonly used. Assessment of age, period, and cohort effects in the cross-sectional design is limited, however, because the effects of one can confound the effects of the others. For instance, in a cross-sectional dataset in which the prevalence of an outcome is observed to be increasing across age, we cannot tell whether this increase is a true age effect or a pure cohort effect (since, in a cross-sectional study, those of different ages are also in different cohorts). Further, recall may be poorer among older individuals, especially for disorders with young age of onset, creating the appearance of a cohort effect. For example, Simon et al. demonstrated that people with major depression are most likely to report onset within the last five years regardless of their age at the time of the interview. Finally, there may be survivor effects, with those surviving to older ages being healthier and less likely to have a history of substance abuse or other psychiatric problems. Thus, while retrospective cross-sectional studies provide important information on cohort effects, they should be interpreted with caution.

5.3.2 Study design: surveillance data

Data collection of multiple age groups observed in many periods is preferable for teasing apart age, period, and cohort effects, but even in this situation, the independent effects of age, period, and cohort cannot be uniquely estimated in standard modelling procedures. This inability to uniquely estimate effects is due to the co-linearity among the three variables; in fact, they are perfectly co-linear. Given an age X and a period Y, the birth cohort Z is fully determined by Y–X. Thus, attempts to model simultaneously all three as additive covariates renders an unidentifiable model without additional assumptions. An array of methods has been developed with varying underlying assumptions in order to achieve model identification. Most models place some type of constraint on the age, period, and cohort covariates in order to induce a non-linear relation with the outcome. Other methods focus entirely on the second order functions of the underlying linear relations, interpreting only the direction of changes over time without specifically ascribing the underlying slope to period or cohort effects. Still other methods define cohort effects as the interaction between age and period and assess systematic variation by cohort that is non-additive across age–period categories. Circumventing identification issues and providing a reinterpretation of common assumptions about how cohort effects arise. We recommend that researchers interested in analysing data use an age–period–cohort model: (a) develop a comprehensive set of hypotheses about how age, period, and cohort effects could arise for a particular psychiatric disorder; (b) how age, period, and cohort effects operate together (e.g. whether effects are hypothesized to be additive or interactive); (c) conduct extensive graphical analysis of the data to understand the descriptive patterns; and (d) choose the best statistical strategy that corresponds to both hypotheses and graphical data patterns.
5.4 Major depression

Evidence indicates that treatment for depression has become more widespread.\textsuperscript{10,11} Whether this is an effect of increased awareness and attention, or of an increase in prevalence, remains to be determined.

We identified 18 population-based studies in the USA that have examined whether individuals in more recently born cohorts have higher prevalence rates of depression compared with older cohorts. Studies published in the 1980s and 1990s using population-based data sources unanimously reported that younger birth cohorts (the youngest birth cohorts of these studies were typically born in the late 1950s and early 1960s) had a higher lifetime prevalence of major depression compared with older birth cohorts (see Table 5.1). Whereas most of these studies were descriptive, three studies attempted formal age–period–cohort modelling in which cohort effects were estimated independently from age and period effects. Results indicated that younger birth cohorts indeed exhibited higher risk, and that there was also a strong period effect in which the risk of major depression was elevated across the whole population. However, virtually all of these studies were cross-sectional analyses relying on respondent report of lifetime depressive episodes. This approach has been widely criticized,\textsuperscript{6,12–14} as differential mortality and recall create bias in examining lifetime rates of depression among older individuals.

Five studies of birth cohort effects in major depression have been published since 2000, and findings are not as consistent as the earlier studies.

Perhaps the most rigorous examination of the issue has been by Costello et al.'s, meta-analysis of data from more than 60,000 children and adolescents from successive birth cohorts.\textsuperscript{15} No changes in the prevalence of depression or depressive symptoms across birth cohorts were documented, leading the authors to conclude that previous investigations based on retrospective reporting were subject to recall bias. Data from the Sterling County study, a longitudinal study across 40 years of data collection, also documented no change in incident depression across two birth cohorts.\textsuperscript{16}

In summary, whereas early population-based studies in the 1980s and 1990s suggested an increase in the prevalence of depression among younger birth cohorts, more recent investigations have revealed little increase in incidence or prevalence, suggesting that: (a) previous studies may have been subject to recall effects, and (b) the perceived increases in depression may well be a function of increased awareness and ascertainment of cases.

5.5 Alcohol use and alcohol use disorders

Substantial evidence indicates that the prevalence of alcohol use, alcohol disorder, and alcohol-related mortality is sensitive to environmental changes such as increases in taxes, changes in policies and laws governing opening and closing hours, happy hours, and threshold for intoxicated driving.\textsuperscript{17} Furthermore, surveillance data indicate substantial trends over time in use of, and mortality due to, alcohol.

Considerable research has been conducted to assess cohort effects in alcohol use disorders, although few studies have been able to formally model age, period, and cohort effects simultaneously. A recent review synthesized evidence from 12 studies that have examined birth cohort effects on alcohol use disorders.\textsuperscript{18} Most studies have been conducted in the USA, with retrospective assessment from cross-sectional surveys used to assess differences in lifetime alcohol use disorder prevalence by birth cohort. A broad selection of birth cohorts across the 20th century has been included in studies of cohort effects on alcohol use disorders. Of these 12 studies, 10 found evidence of cohort effects in alcohol use disorder prevalence, and nine of those 10 found that the youngest cohorts were at the highest risk. These studies have by and large shown consistent
<table>
<thead>
<tr>
<th>Study</th>
<th>Population/design</th>
<th>Cohort range</th>
<th>Depression diagnosis</th>
<th>Findings</th>
<th>Limitations</th>
<th>Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weissman et al.41</td>
<td>Cross-sectional survey of community sample in New Haven</td>
<td>1881–1963</td>
<td>DSM-III MDE</td>
<td>Increasing 6-month and lifetime prevalence and earlier age of onset in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Gershon et al.42</td>
<td>Cross-sectional survey of relatives of US individuals with psychiatric disorders</td>
<td>&lt;1910–1959</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Lavori et al.43</td>
<td>Cross-sectional survey of relatives of US individuals with psychiatric disorders</td>
<td>1930–1959</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence and younger age of onset in more recently born cohorts; strong period effects observed</td>
<td>Retrospective report</td>
<td>Formal age–period–cohort model</td>
</tr>
<tr>
<td>Wickramaratne et al.44</td>
<td>Cross-sectional survey of five US community samples (ECA)</td>
<td>1905–1965</td>
<td>DSM-III MDE</td>
<td>Increasing lifetime prevalence and younger age of onset in more recently born cohorts; period effects also observed for men</td>
<td>Retrospective report</td>
<td>Formal age–period–cohort model</td>
</tr>
<tr>
<td>Klerman et al.45</td>
<td>Cross-sectional survey of relatives of US individuals with psychiatric disorders</td>
<td>&lt;1915–1955 or later</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Joyce et al.46</td>
<td>Cross-sectional survey of New Zealand community adults</td>
<td>1921–1968</td>
<td>DSM-III MDE</td>
<td>Increasing lifetime prevalence in more recently born cohorts (increasing 12-month prevalence for more recently born cohorts in men only)</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Warshaw et al.47</td>
<td>Cross-sectional survey of relatives of US individuals with psychiatric disorders</td>
<td>1906–1965</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence and younger age of onset in more recently born cohorts; strong period effects observed</td>
<td>Retrospective report</td>
<td>Formal age–period–cohort model</td>
</tr>
</tbody>
</table>
Table 5.1 (continued) Population-based studies of birth cohort effects in depression in the USA

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/design</th>
<th>Cohort range</th>
<th>Depression diagnosis</th>
<th>Findings</th>
<th>Limitations</th>
<th>Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ryan et al.⁴⁸</td>
<td>Cross-sectional survey of siblings of children with MDD</td>
<td>(Range not given; mean: 1970; SD: 6)</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Burke et al.⁴⁹</td>
<td>Cross-sectional survey of five US community samples (ECA)</td>
<td>&lt;1917–1966</td>
<td>DSM-III MDE</td>
<td>Increasing lifetime prevalence and younger age of onset in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Cross-National Collaborative Group⁵⁰</td>
<td>Pooled analysis of nine cross-sectional North American general population and family studies (~39,000 subjects)</td>
<td>&lt;1905–1995 or later</td>
<td>DSM-III MDE</td>
<td>Increasing lifetime prevalence in more recently born cohorts, with variation by country</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Lewinsohn et al.⁵¹</td>
<td>Three cross-sectional surveys of adults and one of adolescents</td>
<td>1900–1972</td>
<td>DSM-III-R MDD</td>
<td>Increasing lifetime prevalence in more recently born cohorts</td>
<td>Retrospective report</td>
<td>Convergent findings across different populations</td>
</tr>
<tr>
<td>Leon et al.⁵²</td>
<td>Cross-sectional survey of relatives of US individuals with psychiatric disorders</td>
<td>1915–1954 and later</td>
<td>RDC unipolar depression</td>
<td>Increasing lifetime prevalence in more recently born cohorts for men; peak in prevalence for 1945–1954 cohort for women</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Kessler et al.⁵³</td>
<td>Cross-sectional survey of US general population (NCS)</td>
<td>1936–1975</td>
<td>DSM-III-R MDE</td>
<td>Increasing lifetime prevalence in more recently born cohorts</td>
<td>Retrospective report</td>
<td></td>
</tr>
<tr>
<td>Murphy et al.¹⁶</td>
<td>40-year longitudinal community study</td>
<td>1952 and 1970</td>
<td>Not specified</td>
<td>No difference in incidence between two cohorts; current prevalence higher among younger women</td>
<td>Longitudinal</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Population/design</td>
<td>Cohort range</td>
<td>Depression diagnosis</td>
<td>Findings</td>
<td>Limitations</td>
<td>Strengths</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Kasen et al.</td>
<td>20-year longitudinal study</td>
<td>1928–1945 vs 1945–</td>
<td>Depression index</td>
<td>Overall increases in depression in younger birth cohorts, but more declines with age compared with older birth cohorts</td>
<td>Longitudinal</td>
<td></td>
</tr>
<tr>
<td>Costello et al</td>
<td>Meta-analysis of epidemiological studies comprising 60,000 children (up to age 18 years)</td>
<td>1965–1996</td>
<td>ICD-9, -10, or DSM-III, III-R, or IV MDD or MDE in the past week, month or year</td>
<td>No effect of year of birth on depression</td>
<td>Large pooled sample; current diagnoses</td>
<td></td>
</tr>
</tbody>
</table>

DSM, Diagnostic and Statistical Manual; MDE, major depressive episode; RDC, Research Diagnostic Criteria; MDD, major depressive disorder; ECA, epidemiological catchment area; NCS (-R) National Comorbidity Survey (Replication); NESARC, National Epidemiologic Survey on Alcohol and Related Conditions.
and monotonic increases in the prevalence of alcohol use disorders by birth cohort, with cohorts born in the latter half of the 20th century at higher risk than those born in the early decades of the 1900s.

Interestingly, age-period-cohort modelling of alcohol consumption and high risk drinking patterns, independently of alcohol use disorders, do not show similar trends. Kerr et al. documented in sequentially conducted cross-sectional samples over 26 years that two simultaneous processes are operative in alcohol consumption trends. Whereas cohorts born in the late 1970s and early 1980s consume alcohol less frequently and have a lower overall consumption mean compared with previously born cohorts, heavy episodic drinking is more prevalent. If true, this suggests that increases in risky patterns of alcohol consumption may portend greater drinking consequences for these cohorts later in life. However, other national studies have not found similar results, and instead have documented that heavy episodic drinking is either decreasing in cohorts born in the 1980s or that period effects explain trends over time in heavy episodic drinking rather than cohort effects. The underlying mechanisms through which age, period, and cohort effects arise in alcohol use and related disorders is unclear, although studies are beginning to test specific hypotheses about the role of various environmental factors. Keyes et al. recently documented the alcohol-related social norms of birth cohorts in the 1960s through the 1990s using nationally representative yearly surveys of adolescents. Investigators found that these cohort-specific social norms are strongly associated with individual risk of binge drinking and increases in alcohol-using occasions, independently of time-period-specific social norms as well as the adolescents' personal attitude toward use. This suggests that one driver of birth cohort effects in alcohol use is the amount of social sanctioning against behaviour from peers, which changes across time. Evidence also indicates that women in more recently born cohorts are increasing drinking at a faster rate than men. Given that social norms for women have changed enormously over the past several decades, this accumulating research underscores the notion that social norms play an important role in shaping alcohol use at the population level.

5.6 Autism spectrum disorders

Diagnoses of autism and autism spectrum disorders have increased in the past two decades in the USA as well as in many other western countries. The reasons for these increases remain controversial, and are thought to derive from some combination of increased awareness, changes in diagnoses and diagnostic practices, and an increase in the population prevalence of individual-level risk.

Autism diagnosis is strongly related to child age, with diagnoses most likely to occur between the ages of 3 and 5 years, although the average age of diagnosis has been decreasing as early markers of autism symptoms are increasingly recognized. Age effects, however, are unlikely to be driving the increased incidence of autism diagnosis as age-adjusted rates still show a marked increase over time. This leaves the possibility of some combination of period and cohort effects as driving the increase in autism diagnoses.

Observations of period and cohort effects have different implications for the aetiology of the increases, with period effects suggesting that widespread environmental factors that influence all age groups at a particular time are responsible for the increase, whereas cohort effects indicate that environmental influences that affect specific age groups at a particular time should be considered. Given the relatively constrained age range over which autism diagnoses typically occur, however, period effects are unlikely drivers of diagnostic increases. Environmental factors would
need to have the same effect on autism diagnoses among children of all ages in order for a period effect to explain the increases in autism incidence, which is unlikely based on the developmental epidemiology of autism.

Indeed, evidence to date is strongly suggestive of powerful cohort effects in the incidence and prevalence of autism.\textsuperscript{35–35} For example, data from the Centers for Disease Control and Prevention funded Autism and Developmental Disabilities Monitoring Network indicate a more than twofold increase in the prevalence of autism spectrum disorders among eight-year-olds in the USA born in 1998 compared to 1994.\textsuperscript{36} (with substantial variation in the size of the increase across states.\textsuperscript{34} and more recent data indicate that the prevalence continues to increase.\textsuperscript{26} Population-based data from the California Department of Developmental Services (DDS) and Denmark indicate consistent increases in the incidence of autism diagnoses across age for each successively younger birth cohort from approximately 1990 onward.\textsuperscript{32,33}

California DDS data also indicate that cohort effects in autism diagnosis incidence are stronger in children who exhibit higher functioning on social interaction, language, and communication compared to those with lower functioning (Figure 5.1).\textsuperscript{37} Although the odds of an autism diagnosis among those exhibiting low functioning were four times higher in the 2002 cohort than in the 1992 cohort, the odds of an autism diagnosis among those exhibiting high functioning were almost 15 times higher.

Cohort effects could arise in autism diagnoses through a number of potential mechanisms. For example, if the prevalence of a risk factor acting at conception exhibits change over time (e.g. paternal age\textsuperscript{38}), then each successively younger cohort will be differentially exposed, manifesting as a cohort effect (e.g. each successively younger cohort has older fathers). Alternatively, an exposure introduced into the population as a whole could differentially affect autism incidence depending on age of exposure. This would also manifest as a cohort effect. As the prevalence of autism diagnoses is continuing to increase, continued surveillance of period and cohort effects in autism spectrum disorders will continue to be an important public health activity.

![Graph showing odds ratio by birth cohort and functioning level.](image)

**Figure 5.1** Cohort effects in autism diagnosis in California from 1994 to 2005 by child's functioning at the time of diagnosis.
5.7 Other disorders

This chapter discusses evidence for age, period, and cohort effects for psychiatric disorders with a well-documented evidence base regarding these effects. Bipolar disorder and attention deficit/hyperactivity disorder (ADHD) have also demonstrated increases in childhood diagnoses, but limited investigations have been made into possible age, period, and cohort effects. The rapid changes in trends over time suggest the need for continued surveillance and age-period-cohort modelling.

5.8 Summary

Assessment of age, period, and cohort effects is vital for a life course approach to psychiatric epidemiology. Variation in the incidence and prevalence of psychiatric disorders by birth cohort, independent of age and period, suggests that the social structures in which individuals are embedded during important developmental time-periods can influence the risk of psychiatric disorder onset for groups of individuals exposed to those structures. Published research supports a role for birth cohort effects in the incidence and prevalence of alcohol use disorders. Whereas cohort effects in major depression have been found, the existence of cohort effects remains controversial due to data suggesting that a retrospective reporting bias may explain a large part of these results. Autism diagnoses have robust evidence for cohort effect. The increase in the prevalence of many psychiatric disorders during childhood, adolescence, and young adulthood may portend greater consequences in numerous domains of social, medical, and psychiatric morbidity throughout the next several decades as these young cohorts reach middle and old age.

Acknowledgements

This research was supported in part by support from New York State Psychiatric Institute and the Department of Epidemiology at the Mailman School of Public Health (K. Keyes).

References


49 Burke KC, Burke JD, Jr, Rae DS, Regier DA. Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. *Arch Gen Psychiatry* 1991;48(9):789–795.


56 Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry* 2005;62(10):1097–1106.
Implementation of Age-Specific Services for Transition-Age Youths in California

Victoria D. Ojeda, Ph.D., M.P.H., Sarah P. Hiller, M.P.I.A., Samantha Hurst, Ph.D., Nev Jones, Ph.D., Sara McMenamin, Ph.D., James Burgdorf, Ph.D., Todd P. Gilmer, Ph.D.

Objectives: This study examined the implementation of age-specific services for transition-age youths in California under the Mental Health Services Act (MHSA).

Methods: This study employed a sequential, exploratory mixed-methods design. Qualitative interviews with 39 mental health service area administrators in California were analyzed to develop an understanding of how the MHSA has facilitated the development of youth-specific programs or services. A quantitative survey of 180 youth-focused programs was also used to describe the range of services that were implemented, the use of evidence-based and promising practices, and the role of youths in the design, planning, delivery, and evaluation of services.

Results: Administrators described the MHSA as providing a programmatic focus and financial support for youth-specific services, outlining a stakeholder process to create buy-in and develop a vision for services, and emphasizing the role of youths in service delivery and planning. Youth-specific programs implemented a diverse array of services, including general medical care; employment and education support; housing placement and support; and family, mentoring, and social support. Programs described implementing evidence-based and promising practices and involving youths in service planning, implementation, or quality improvement activities.

Conclusions: The MHSA has had a substantial impact on the landscape of youth-specific services in California by expanding both the number of programs and the diversity in types of services and by promoting the engagement of youths in the planning and delivery of services. Additional efforts are necessary to determine the extent to which youth-specific services yield greater improvements in youths’ outcomes compared with services designed for adults.

Psychiatric Services in Advance (doi: 10.1176/appi.ps.201500084)

Young adulthood is a critical developmental period, during which decisions made in areas such as education, employment, and parenting have lasting effects on lifetime trajectories (1). Economic restructuring, including increased costs of college and a decline in well-compensated entry-level employment, has increased the demands on youths and has changed and extended their pathways to independence (1). Successful transitions to adulthood are less likely among youths with serious mental illness, who, compared with their peers without mental illness, have lower rates of education and employment and higher rates of poverty, unplanned pregnancy, substance use disorders, homelessness, and criminal justice involvement (1–6). The challenges inherent in the transition to adulthood are often more difficult among foster care youths because of their emancipation, lack of natural mentors, justice system involvement, and needs for mental health services related to life transitions that are not adequately met by a mental health service system that is bifurcated for adults and children (1,7–9). Concerns about the mental health system’s ability to engage and retain youths in treatment have prompted calls for services that are age specific and developmentally appropriate (1,10,11).

On November 2, 2004, California voters approved Proposition 63, which was signed into law as the Mental Health Services Act (MHSA). The MHSA applies a tax of 1% on incomes over $1 million to fund public mental health services (12). The MHSA provides new funding streams for specific types of services, such as full-service partnerships, and for priority populations identified as being underserved by the public mental health system (13). Transition-age youths are designated as a priority population under the MHSA, and public mental health agencies have used the opportunity to develop services tailored to this population (14,15).

Consistent with prior efforts focused on reforming the delivery of mental health care in California, the MHSA provides broad policy guidance but relied on local mental health agencies to design and implement new programs (16). As a result, we lack a clear understanding of how the MHSA has facilitated the development of youth-specific services, as well as the scale and scope of programs that have been
implemented for youths. Funding provided under the MHSA can be used only to expand mental health services and not to supplant existing state or county funds used to provide mental health services (17). This study aimed to identify specific policies and processes by which the MSHA has facilitated the development of youth-specific services and to describe the range of services that have been implemented for transition-age youth in California under the MHSA. This study also examined the use of evidence-based or promising practices and the role of youths in the design, planning, delivery, and evaluation of services (18).

METHODS

Study Design and Sample
This 2014 study employed a sequential, exploratory mixed-methods design (19). Qualitative, semi-structured interviews with administrators in California’s mental health service areas were used to assess how the MHSA has facilitated the development of youth-specific programs. Administrators were individuals who were involved in the development and oversight of programs for transition-age youths. A quantitative survey administered to youth service providers was used to describe the range of services that were implemented for transition-age youths (20). Providers were either program managers or agency or center directors. The qualitative and quantitative data were connected to provide complementarity, whereby the overarching policy processes of the MHSA as described by the interview respondents was compared with the actual services provided by youth-specific mental health programs in California.

California has 59 independent public mental health service areas: 57 areas are individual counties, two are separate city-based areas within counties, and one is a combination of two counties. Queries to mental health administrators and a review of MHSA service planning documents showed that 48 service areas had implemented youth-specific programs, including stand-alone youth-specific programs and youth-specific services embedded within larger non-youth-specific programs. Administrators were invited to participate in a semi-structured phone interview and also asked for contact information for directors of programs providing youth-specific services who could respond to an online survey. Interviews were conducted with administrators in 39 service areas (81% response rate), and 180 out of 298 programs completed the program survey (60% response rate). [A table presenting detailed response rates by county is included in an online supplement to this article.]

Evaluation Advisory Group
This study was informed by an advisory group of three mental health administrators with expertise in implementing youth-specific programs, three youth-specific program directors, two program evaluators with context expertise, and two youths with both lived experience and expertise in program evaluation. Community involvement by key stakeholders was considered critical to ensure that the instruments developed for collecting data and the analyses conducted were responsive to the needs of administrators, program directors, and clients. The advisory group provided input on the research design, interview guide, survey questions, and interpretation of findings. The University of California, San Diego, Human Subjects Research Protections Program approved this study.

Data Collection and Analysis
Interviews were conducted with administrators representing 39 mental health service areas to discuss their perception of the role of the MHSA in the provision of services for transition-age youths. The interviews were conducted via telephone or in person in a setting of their choice (for example, in their office). Interviews were audio recorded and transcribed verbatim. Transcripts were read and organized using Dedoose qualitative analysis online software. Directed and conventional content analysis techniques were used to assign codes on the basis of a priori themes derived from the interview guide and on emergent themes raised by respondents (21). The qualitative analyses focused on responses related to the overarching a priori theme of MHSA’s facilitation of the development of youth-specific services. The transcripts were independently coded by two authors (SPH and SH) with expertise in medical anthropology and public health, under the supervision of the principal investigators (VDO and TPG). Disagreements in assignment or description of codes were resolved by team discussion. Segments of narrative were coded for descriptive content domains that emerged from the responses and then analyzed by using matrix analysis (22). The matrices were constructed for each descriptive content domain by creating a table to display the mental health service areas in the first column and the emergent themes in the column headings. Coded interview data were displayed by theme in individual vertical cells matching each of the service areas represented in the first column. This process allowed team members to view the body of data holistically, noting systematic similarities as well as trends or differences in the constructs of information across the interviews.

Youth-specific programs were requested to respond to an online self-administered survey. The survey collected quantitative data on program characteristics, the types of services provided for transition-age youths, the use of evidence-based practices, and the involvement of youths in service planning and governance. Lists of services were developed on the basis of a review of MHSA service planning documents, the authors’ knowledge of general and youth-specific mental health services, and input from the evaluation advisory group. This list of evidence-based and promising practices for youths was also informed by a review of the literature and various data repositories of evidence-based treatments.

Answers to the survey were multiple choice and included a residual “other” category in which respondents could enter additional textual information. Categories were added when a sufficient number of programs had the same “other” response. Programs were asked to identify any important remaining gaps in services for youths. This question requested
a text response, and answers were coded according to the service categories provided earlier in the survey. The survey was administered using Qualtrics online software. Weekly reminders were sent to program directors who had not initiated or completed their surveys. The survey data were reviewed and cleaned as responses were received, and clarifying queries were made to the respondents as necessary.

Descriptive analyses were used to characterize the scale and scope of youth-specific services and level of youth involvement in various activities. Regression analyses were used to compare services across counties by level of urbanization. Counties were placed into one of three categories on the basis of the size of their metropolitan statistical areas (MSAs) by using the 2013 National Center for Health Statistics urban-rural classification scheme (23): large central metro or large fringe metro (counties in MSAs of one million or more), medium or small metro (counties in MSAs of less than one million), or micropolitan or noncore (nonmetropolitan areas). A series of negative binomial regression models estimated the relationship between urbanization and the number of youth-specific services at the program level, controlling for funding type.

RESULTS

MHSA's Role in Facilitating Development of Age-Specific Services

Table 1 presents three emergent themes and illustrative quotes. The overarching a priori theme was the role of the MHSA in facilitating the development of youth-specific services. The first emergent theme was the MHSA's role in providing both a programmatic focus and a funding stream for youth-specific services. Most administrators described the MHSA as the sole reason for the existence of youth-specific services. Two administrators reported having small existing programs funded by federal grants. Many administrators described developing an array of programs or a youth-focused system of care. Common elements of the MHSA-funded system of care included permanent supportive housing and transitional housing, early intervention programs, youth-specific outpatient programs, drop-in counseling centers, employment support, life skills and vocational training, and community-embedded outreach services based in local high schools. Administrators credited the MHSA as spreading awareness of the need for age-specific services and promoting broader buy-in.

A second emergent theme considered the role of the stakeholder process outlined by the MHSA to support the identification and implementation of youth-specific services. The stakeholder process helped to create buy-in and to develop a vision for youth-specific services. A third emergent theme focused on the MHSA's role in emphasizing the role of youths in the delivery and development of youth-specific services. Youths were engaged as peer mentors, facilitators, and advisors to improve the effectiveness and relevance of services. Several county programs engaged youths in outreach

### TABLE 1. Themes and illustrative quotes from interviews in 2014 with administrators representing 39 mental health service areas about the role of the Mental Health Services Act (MHSA) in facilitating development of youth-specific services

<table>
<thead>
<tr>
<th>Theme</th>
<th>Illustrative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHSA provided the programmatic focus and funding stream for youth-specific services</td>
<td>“I think the MHSA has provided us a very unique opportunity to be able to develop an array of services, from prevention to intervention, in areas where prior to the MHSA we did not have the funding or the really organized stakeholder process to move forward. “ Virtually all of our services are funded under the MHSA. Our system of care we’ve developed from MHSA. And we did our best to create a continuum of care for transition-age youths struggling with a severe mental illness or being identified as having first symptoms of a serious mental illness. It’s really allowed us to expand. We’ve moved from one program to now having 15, specifically for the transition-age youths. “ Without MHSA our programs would not be in existence. The way that the MHSA laid out the priority populations by age and grouping them, I think, certainly helped facilitate that process.”</td>
</tr>
<tr>
<td>Outlining a stakeholder process to create buy-in and develop a vision for youth-specific services</td>
<td>“[The process] has really initiated or “refined” the community to speak to the need of programming for our transition-age youths. ” [We] utilized the stakeholder process to identify transition-age youths as an uninsured and underserved population. And we really developed a very, very rich vision for a continuum of age-specific programming.” “I think [the process] has led people to recognize that transition-age youths are a population with unique needs and interests. . . . It’s really revolutionized our whole way of conceptualizing the continuum of services, from childhood to adulthood.”</td>
</tr>
<tr>
<td>Engaging youths in the delivery and development of services</td>
<td>“We use [peer] mentors for outreach and engagement on all levels of programming.” “We hired the very first two peer and family advocates, and we started listening to them, and everyone was exposed to them throughout the system. It sort of opened people’s ears up.” “We involve transition-age youths in the planning, implementation, and evaluation of programs. It’s sort of changed our culture to some degree on how we do things.” “We have been trying to create credible images of what a transition-age youth is, and so it’s really about how do we use youths in our system to make a difference? And so we use youths where we believe that youth voices are important, and we want to have a place for them to be able to use their voice. We invite them to our cultural competency committee. We have youths on our mental health advisory board. We also use them like we’re doing for our bullying campaign—we’re using them as facilitators. But the reason is because they thrive when they can use their voice and their voice is heard. We need to keep our services relevant and up to date and we use the youth voice to do that.”</td>
</tr>
</tbody>
</table>
to high schools in campaigns to address bullying, stigma, and suicide.

Implementation of Age-Specific Services Under the MHSA

Youth-specific services were implemented in 48 of California’s 59 mental health service areas (81%). Only 11 (52%) of nonmetropolitan (that is, micropolitan or noncore) areas implemented youth-specific programs, compared with 19 (95%) of the medium and small metro areas and 18 (100%) of large central metro or large fringe metro areas.

Table 2 presents data on the funding streams employed and priority populations targeted by youth-specific programs in California. Notably, 164 programs (91%) received MHSA funding and thus represented new programs or services; 60 programs (33%) reported receiving funding from Medicaid through the Early and Periodic Screening, Diagnosis, and Treatment program, indicating that youth-specific programs leveraged multiple funding sources.

Table 2. Characteristics of 180 programs providing youth-specific services in California in 2014

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>%</th>
<th>Characteristic</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Health Services Act (MHSA) funding streams employed to support services</td>
<td>164</td>
<td>91</td>
<td>Programs targeting specific priority populations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-service partnerships (FSPs): supportive housing programs that “do whatever it takes” to improve residential stability and engage clients in services</td>
<td>67</td>
<td>37</td>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community services and supports: non-FSP, community-based behavioral health services</td>
<td>44</td>
<td>24</td>
<td>Transition-age youths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevention and early intervention: services designed to provide early intervention or to engage clients before development of serious mental illness or emotional disturbance</td>
<td>89</td>
<td>49</td>
<td>With serious mental illness</td>
<td>126</td>
<td>70</td>
</tr>
<tr>
<td>Other MHSA funding</td>
<td>38</td>
<td>21</td>
<td>With co-occurring mental and substance use disorders</td>
<td>126</td>
<td>70</td>
</tr>
<tr>
<td>Non-MHSA funding</td>
<td>72</td>
<td>40</td>
<td>Residential setting of transition-age youths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early and Periodic Screening, Diagnosis, and Treatment: Medicaid-funded behavioral health services</td>
<td>60</td>
<td>33</td>
<td>Homeless or at risk of becoming homeless</td>
<td>118</td>
<td>66</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>7</td>
<td>Aging out of the foster care system</td>
<td>106</td>
<td>59</td>
</tr>
<tr>
<td>Annual caseload</td>
<td></td>
<td></td>
<td>Exiting the juvenile or criminal justice system</td>
<td>100</td>
<td>56</td>
</tr>
<tr>
<td>1–49</td>
<td>43</td>
<td>24</td>
<td>In institutional or residential care</td>
<td>61</td>
<td>34</td>
</tr>
<tr>
<td>50–99</td>
<td>47</td>
<td>26</td>
<td>Race-ethnicity of transition-age youths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100–249</td>
<td>46</td>
<td>25</td>
<td>Latino</td>
<td>92</td>
<td>51</td>
</tr>
<tr>
<td>250–4,500</td>
<td>44</td>
<td>25</td>
<td>African American/black</td>
<td>79</td>
<td>44</td>
</tr>
<tr>
<td>Age of clients (M±SD)</td>
<td></td>
<td></td>
<td>Asian American</td>
<td>61</td>
<td>34</td>
</tr>
<tr>
<td>Minimum</td>
<td>16±2</td>
<td></td>
<td>Native American/Alaska Native</td>
<td>57</td>
<td>32</td>
</tr>
<tr>
<td>Maximum</td>
<td>24±2</td>
<td></td>
<td>Pacific Islander</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>8</td>
<td>Native Hawaiian</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>Other characteristic of transition-age youths</td>
<td>Other</td>
<td>8</td>
<td>Undocumented immigrant</td>
<td>68</td>
<td>38</td>
</tr>
<tr>
<td>Lesbian, gay, bisexual, transgender, or queer</td>
<td>88</td>
<td>49</td>
<td>Veteran</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>Parenting or pregnant</td>
<td>77</td>
<td>43</td>
<td>Other</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Undocumented immigrant</td>
<td>68</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Youth-specific programs also focused on services that help youths build their human capital and strengthen their financial skills, including services to support education and employment goals as well as financial and benefits management services. Approximately one-half of the programs provided coordination with secondary and postsecondary schools, and about one-third offered employment counseling or supported employment. Approximately one-half of the programs provided housing placement and support; most common was transitional or time-limited housing and independent housing.

Family counseling and psychoeducation were commonly offered, as were parenting skills for pregnant or parenting youths. Mentoring and peer support services included peer mentoring, supporting natural mentors, and peer-led
support groups. Other social support services included community integration, recreational activities, and social skills training.

Regression analysis indicated that nonmetropolitan counties provided 1.9 fewer types of behavioral health services and 1.3 more types of peer services, compared with large central metro or large fringe metro counties (p<.05 each, data not shown). No significant differences were found by residential setting in other service types.

Table 4 presents data showing that most programs reported use of one or more evidence-based or promising practices. The most common evidence-based model was early intervention for psychosis. The most common process or approach was motivational interviewing, and the most common manualized therapy was trauma-informed cognitive-behavioral therapy.

As shown in Table 5, about two-thirds of programs involved youths in service planning, and nearly half involved youths as members of their advisory group, on planning or implementation committees, or in evaluation or quality improvement activities. Few programs included youths as members of their governing bodies.

Programs were queried about any remaining gaps in services for youths. The most frequently mentioned gaps included youth-specific housing and homelessness services, including needs for more independent and supportive housing, transitional housing, and youth-specific emergency overnight shelters (N=63, 35%). Gaps in youth-specific behavioral health services included outpatient and crisis residential services (N=42, 23%), and gaps in youth-specific employment support included supported employment, job placement, and job development (N=18, 10%).

### TABLE 3. Services offered in 2014 by 180 youth-specific programs in California

<table>
<thead>
<tr>
<th>Service</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral health service</td>
<td>164</td>
<td>91</td>
</tr>
<tr>
<td>Assessment</td>
<td>149</td>
<td>83</td>
</tr>
<tr>
<td>Case management</td>
<td>155</td>
<td>86</td>
</tr>
<tr>
<td>Crisis intervention</td>
<td>146</td>
<td>81</td>
</tr>
<tr>
<td>Intensive case management</td>
<td>103</td>
<td>57</td>
</tr>
<tr>
<td>Medication management</td>
<td>142</td>
<td>79</td>
</tr>
<tr>
<td>Individual or group therapy</td>
<td>117</td>
<td>65</td>
</tr>
<tr>
<td>Mental health rehabilitation services</td>
<td>142</td>
<td>79</td>
</tr>
<tr>
<td>Substance abuse treatment</td>
<td>80</td>
<td>44</td>
</tr>
<tr>
<td>General medical services</td>
<td>113</td>
<td>63</td>
</tr>
<tr>
<td>Coordination with general medical providers</td>
<td>91</td>
<td>51</td>
</tr>
<tr>
<td>Family planning</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>Health education for prevention of HIV</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>Health education for prevention of sexually transmitted infections</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>General medical care</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Physical wellness programs</td>
<td>32</td>
<td>18</td>
</tr>
<tr>
<td>Screening and assessment</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Screening for alcohol abuse</td>
<td>38</td>
<td>21</td>
</tr>
<tr>
<td>Testing for HIV infection</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Testing for sexually transmitted infections</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Education, employment, benefits, and financial services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education support</td>
<td>112</td>
<td>62</td>
</tr>
<tr>
<td>Coordination with secondary and postsecondary schools</td>
<td>99</td>
<td>55</td>
</tr>
<tr>
<td>Educational counseling</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>Educational testing and assessment</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>Supported education</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>Employment support</td>
<td>98</td>
<td>54</td>
</tr>
<tr>
<td>Employment and vocational testing and assessment</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>Employment counseling and placement</td>
<td>62</td>
<td>34</td>
</tr>
<tr>
<td>Supported employment</td>
<td>69</td>
<td>38</td>
</tr>
<tr>
<td>Transitional employment</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>Benefits management</td>
<td>102</td>
<td>57</td>
</tr>
<tr>
<td>Benefits advocacy or enrollment</td>
<td>98</td>
<td>54</td>
</tr>
<tr>
<td>Benefits coordination or management</td>
<td>57</td>
<td>32</td>
</tr>
<tr>
<td>Legal assistance</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>Financial services</td>
<td>101</td>
<td>56</td>
</tr>
<tr>
<td>Debt restructuring</td>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>101</td>
<td>56</td>
</tr>
<tr>
<td>Housing and basic services</td>
<td>87</td>
<td>48</td>
</tr>
<tr>
<td>Housing placement and support</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>Congregate housing</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>Crisis residential</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>Emergency shelter or respite</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>Independent housing</td>
<td>55</td>
<td>31</td>
</tr>
<tr>
<td>Transitional or time-limited housing</td>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>Vouchers or rental subsidies</td>
<td>113</td>
<td>63</td>
</tr>
<tr>
<td>Basic services</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Clothing</td>
<td>65</td>
<td>36</td>
</tr>
<tr>
<td>Communication services</td>
<td>47</td>
<td>26</td>
</tr>
<tr>
<td>Laundry services or tokens</td>
<td>56</td>
<td>31</td>
</tr>
<tr>
<td>Meals or vouchers</td>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>Showers</td>
<td>106</td>
<td>59</td>
</tr>
<tr>
<td>Family, mentoring, and social services</td>
<td>144</td>
<td>80</td>
</tr>
<tr>
<td>Family services</td>
<td>114</td>
<td>63</td>
</tr>
<tr>
<td>Family counseling</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>Family events</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>Family reunification services</td>
<td>89</td>
<td>49</td>
</tr>
<tr>
<td>Family psychoeducation</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>Multifamily group therapy</td>
<td>81</td>
<td>45</td>
</tr>
<tr>
<td>Parenting skills for pregnant or parenting transition-age youth</td>
<td>108</td>
<td>60</td>
</tr>
<tr>
<td>Mentoring support</td>
<td>80</td>
<td>44</td>
</tr>
<tr>
<td>Peer mentoring</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>Professional mentoring</td>
<td>59</td>
<td>33</td>
</tr>
<tr>
<td>Supporting natural mentors</td>
<td>97</td>
<td>54</td>
</tr>
<tr>
<td>Peer support</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>Peer-led counseling</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>Peer-led drop-in center or day program</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>Peer-led education</td>
<td>79</td>
<td>44</td>
</tr>
<tr>
<td>Peer-led support groups</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Peer-led warm line</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Peer-run crisis intervention</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Peer-run crisis residential</td>
<td>96</td>
<td>53</td>
</tr>
<tr>
<td>Social support services</td>
<td>112</td>
<td>62</td>
</tr>
<tr>
<td>Community integration and inclusion support</td>
<td>140</td>
<td>78</td>
</tr>
</tbody>
</table>
The majority of programs included in this study. In contrast, most services investigated previously are offered in less than 10% reported offering these services statewide (24). In offering any single type of youth-specific services. A few administrators who described having pre-existing, federally funded programs also credit the MHSA as the primary catalyst for implementation of youth-specific services. Administrators described the MHSA funding, and the MHSA has provisions against using the funds to supplant existing state or county funding for mental health services. Administrators described the MHSA as providing a programmatic focus and financial support for youth-specific services, outlining a stakeholder process to create buy-in and develop a vision for services and emphasizing the role of youths in service delivery and planning. Youth-specific programs implemented a diverse array of services, including general medical care; support of education and employment; housing placement and support; and family, mentoring, and social support. Programs largely described implementing evidence-based practices and involving youths in service planning, implementation, and quality improvement activities.

The youth-specific services described here are largely the result of an expansion in services under the MHSA. An overwhelming majority of programs reported receiving MHSA funding, and the MHSA has provisions against using the funds to supplant existing state or county funding for mental health services. Administrators described the MHSA as the primary catalyst for implementation of youth-specific services. A few administrators who described having pre-existing, federally funded programs also credit the MHSA as supporting the continuation and expansion of these programs. The landscape of youth-specific services in California is substantially larger than what was offered nationally almost 15 years ago, when fewer than 20% of states reported offering any single type of youth-specific service and fewer than 10% reported offering these services statewide (24). In contrast, most services investigated previously are offered in the majority of programs included in this study.

### DISCUSSION

This study used mixed methods to identify specific policies and processes by which the MHSA facilitated the development of youth-specific services and to describe the range of services that have been implemented for transition-age youths in California under the MHSA. Administrators of mental health service areas described the MHSA as providing a programmatic focus and financial support for youth-specific services, outlining a stakeholder process to create buy-in and develop a vision for services and emphasizing the role of youths in service delivery and planning. Youth-specific programs implemented a diverse array of services, including general medical care; support of education and employment; housing placement and support; and family, mentoring, and social support. Programs largely described implementing evidence-based practices and involving youths in service planning, implementation, and quality improvement activities.

The youth-specific services described here are largely the result of an expansion in services under the MHSA. An overwhelming majority of programs reported receiving MHSA funding, and the MHSA has provisions against using the funds to supplant existing state or county funding for mental health services. Administrators described the MHSA as the primary catalyst for implementation of youth-specific services. A few administrators who described having pre-existing, federally funded programs also credit the MHSA as supporting the continuation and expansion of these programs. The landscape of youth-specific services in California is substantially larger than what was offered nationally almost 15 years ago, when fewer than 20% of states reported offering any single type of youth-specific service and fewer than 10% reported offering these services statewide (24). In contrast, most services investigated previously are offered in the majority of programs included in this study.

### TABLE 4. Use of evidence-based and promising practices in 2014 by 180 youth-specific programs in California

<table>
<thead>
<tr>
<th>Practice</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of any evidence-based or promising practice</td>
<td>160</td>
<td>89</td>
</tr>
<tr>
<td>Model</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>Assertive community treatment</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Early intervention for psychosis</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Individual Placement and Support</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Permanent supportive housing</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Process or approach</td>
<td>128</td>
<td>71</td>
</tr>
<tr>
<td>Managing and adapting practice</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Motivational interviewing</td>
<td>107</td>
<td>59</td>
</tr>
<tr>
<td>Transition to Independence Process</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>Manualized therapy</td>
<td>136</td>
<td>76</td>
</tr>
<tr>
<td>Cognitive-behavioral therapy for psychosis</td>
<td>58</td>
<td>32</td>
</tr>
<tr>
<td>Dialectical behavior therapy</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>Integrated dual diagnosis treatment</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>Seeking Safety</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>Trauma-informed cognitive-behavioral therapy</td>
<td>91</td>
<td>51</td>
</tr>
<tr>
<td>Other manualized therapy</td>
<td>29</td>
<td>16</td>
</tr>
</tbody>
</table>

### TABLE 5. Involvement of transition-age youths in program planning and governance in 180 youth-specific programs in California in 2014

<table>
<thead>
<tr>
<th>Activity</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any involvement</td>
<td>144</td>
<td>80</td>
</tr>
<tr>
<td>Involved in service planning</td>
<td>118</td>
<td>66</td>
</tr>
<tr>
<td>Members of an advisory group or council</td>
<td>86</td>
<td>48</td>
</tr>
<tr>
<td>Involved in evaluation or quality improvement activities</td>
<td>76</td>
<td>42</td>
</tr>
<tr>
<td>Members of program’s governing body</td>
<td>24</td>
<td>13</td>
</tr>
</tbody>
</table>

A majority of youth-specific programs implemented evidence-based or promising practices. However, it is important to note that most evidence-based practices do not have evidence of efficacy specifically for youths. Notable exceptions include supportive age-specific evidence for Individual Placement and Support (both standard and as adapted for early intervention in psychosis), the Transition to Independence Process, and motivational interviewing (25–30). Given significant developmental differences between youths and mature adults, it should not be assumed that practices that have been shown to be effective for adults are similarly effective for youths. At least in some cases, evidence-based practices may need significant adaptation to adequately meet the needs of youths (1).

Youths were actively engaged in the design and delivery of services. The perceived benefits of greater youth involvement include programs’ greater cultural competency with respect to youths, more effective outreach to youths in community settings, and a greater relevance of services to youths. Additional research should examine the degree to which youth involvement results in improvements in these areas.

This study had several limitations. Program data were obtained by using a self-administered survey. This is an expeditious approach to obtaining information on a critical array of practices across a wide range of programs, but it lacks some detail in measurement compared with what might be obtained from service utilization data; the depth of understanding or context that might be obtained from site visits is also lacking with this approach. Participation in the survey was voluntary, and not all California programs participated. Most missing responses originated from Los Angeles County. Although the MHSA has provisions against funds being used to supplant existing state or county funding for mental health services, we do not have concrete information on the extent to which this may have occurred. The data on use of evidence-based practices did not include information on fidelity to their respective models. Although this study did not investigate the impact of changing the nature of services for youths on their health or recovery outcomes, previous research demonstrated its impact on improved treatment attendance, and future research should investigate the impact on these important outcomes (14,15). The legislative approach pursued in California can serve as a policy model for other states that seek to improve services for this population.
CONCLUSIONS

The MHSA has had a substantial impact on the landscape of youth-specific services in California by providing a programmatic focus and financing and by outlining a stakeholder process that has created buy-in and a vision for a youth-oriented delivery system. The MHSA has also been influential in promoting the engagement of youths in multiple areas of the service delivery system. Additional efforts are necessary to determine the extent to which the provision of youth-specific services yields greater improvements in outcomes compared with the provision of services that are designed for adults.

AUTHOR AND ARTICLE INFORMATION

Dr. Ojeda and Ms. Hiller are with the Department of Medicine, and Dr. Hurst, Dr. McMenamin, and Dr. Gilmer are with the Department of Family Medicine and Public Health, all at the University of California, San Diego, La Jolla (e-mail: vojeda@ucsd.edu). Dr. Jones is with the Felton Institute, San Francisco. Dr. Burgdorf is with Medimpact, San Diego.

The authors report no financial relationships with commercial interests.

Received February 26, 2015; revisions received October 23 and December 15, 2015; accepted February 1, 2016; published online May 1, 2016.

REFERENCES

1. Investing in the Health and Well-Being of Young Adults. Washington, DC, National Academies Press, 2015
9. Vostanis P: Patients as parents and young people approaching adulthood: how should we manage the interface between mental health services for young people and adults? Current Opinion in Psychiatry 18:449–454, 2005
INTRODUCTION

An appreciation for sex as a biological variable (SABV) in preclinical and clinical research is essential for our understanding of basic mechanisms contributing to risk and resilience, especially in cases where there are known sex differences in phenomenology, natural history, treatment, or severity. Studies focusing on examining sex differences have demonstrated across many levels of analyses and stages of brain development and maturation that males and females can differ significantly. From sex chromosomes that allow for disparities in gene dosage and regulatory mechanisms, to the important role of gonadal hormones across the life span, our appreciation for the unique and highly mechanistic insight that including SABV affords us has gained great attention. This review will provide general oversight as to how to complete studies in this area in preclinical and clinical research (Box 1; Figures 1 and 2). Finally, the role of developmental windows and changes across the life span to the contribution of sex differences in neuropsychiatric and cognitive disorders is highlighted where appropriate and in Figure 3.

Defining Sex and Gender

Although all cells have a sex, designated by the presence and dosage of X or Y chromosomes, which in most cases will be XX (female) or XY (male), gender is a designated societal determinant and therefore traditionally described only in humans (Bachmann and Mussman, 2015). Of particular importance for this discussion, rodents do not have a gender. Discussed below are the primary contributions of gonadal hormones and sex chromosomes in organizing the sexually dimorphic brain, a key foundation throughout life where the internal and external environments can provoke phenotypic sex differences.

Importantly, the sexually dimorphic brain, similar to most sex differences, does not fall into a hard binary readout—but rather is on a continuum or spectrum with each cell and each brain region comprised of varying degrees of ‘male’ and ‘female’ (Hines, 2005; Joel and McCarthy, 2016). This is because the influences from very early neurodevelopmental time points, and perhaps even earlier than fertilization, are complex and multifaceted and frequently depend on the sex chromosome compliment of the individual or the sex of the parent contributing a given gene. Finally, the combination of genetic sex and gonadal hormones alone can produce a plethora of points at which subtle differences promote dramatic trajectory changes in development.

An example of this is from Catherine Dulac’s group in which they investigated the genetics of sex differences in the brain by comparing parent-of-origin allelic gene expression and found that imprinted genes inherited from maternal or paternal origin were dependent on the offspring sex as
Box 1  How to Examine the Estrous Cycle in Rodents

One simple way in which cyclicity can be examined is to collect vaginal smears on the day of testing and used to determine cycle stage in females as described previously (Nelson et al., 1982; Caligioni, 2009). Cycling should be done immediately after or as soon as possible after the behavioral study is completed such that you are not stressing the female before testing and you are obtaining cycle information in as close time period to the behavioral measure as possible. To obtain samples for cytological comparison by vaginal lavage, use warmed saline to carefully triturate a small volume from a transfer pipet placed just at the vaginal opening. Caution here, too vigorous of trituration or placed too far inside the vagina can induce pseudopregnancy, so be gentle. Transfer the fluid to a microscope slide, and allow the sample to dry at room temperature. To identify the estrous cycle stage the female is in, the types of cells present (cornified epithelial, nucleated epithelial, and leukocytes) and ratio of these cells is determined by crystal violet staining of the slides as previously described (Caligioni, 2009). Additional methods are also available for confirmation of estrous cycle stage (Becker et al., 2005).

Figure 1  How to examine menstrual cycle impact in human studies. Depicted is an idealized 28-day cycle, although studies for which menstrual cycle length is considered frequently include women with regular cycles from 25 to 34 days in length from day 1 (first day of menstrual flow) to day 1 (first day of menstrual flow) of the next cycle. If necessary, varying cycle lengths can be modified statistically to conform to a standardized 28-day cycle. However, the most straightforward way to assess changes in outcomes across the cycle or between phases is to target when a given woman is likely to be in the hormonal and/or behavioral state of interest (Epperson et al., 2007; Comasco et al., 2014; Timby et al., 2016). Many women will have several days of spotting before experiencing a full flow. Whether investigators choose to consider any spotting as day 1 or onset of full flow as day 1 varies, but should remain consistent within a given study.

Figure 2  Sex differences in gonadal and pituitary hormones in middle- and late-aged adults. Data depicting general fluctuations in reproductive hormones across the life span from the typical age of late pre-menopause to the late post-menopause in women and age-matched men (Moroz and Verkhoshansky, 1985; Liu et al., 2015). A rapid effect is indicated by two arrows. A slow or moderate effect on gonadal and pituitary hormones is indicated by one arrow.

Well (Gregg et al., 2010). Within the neuroendocrine hypothalamus, this study showed that there was an important interaction at the level of epigenetic modifications in controlling gene expression patterns between parent and offspring sex. At a very molecular level, this is an incredible example of where the same parents, the same environment, and the same gestational experience can yield significantly different outcomes in brain development that may then predict long-term sex differences in disease risk and resilience.

Also, critical to the discussion of sex and gender in the human laboratory is the interaction between an individual’s experiences, based upon society’s concept of their gender, and the developing central nervous system (Borchers and Gershwin, 2012; Springer et al., 2012). Examples of this phenomenon include opportunities for education and enrichment, exposures to particular types of trauma and injury, and access to and uptake of behavioral health treatment. It is well known that education is important to long-term cognitive health and although women in the United States now outnumber men with respect to graduate degrees, women in some communities are not encouraged to seek higher education and overall health status is adversely impacted (Hendi, 2015). Likewise, females are more likely than males to experience sexual abuse during critical periods of brain development and maturation, particularly during peri-adolescence when sub-cortical structures are vulnerable to insult (Felitti et al, 1998). Although gender differences in sports-related concussions vary by type of sport, females show higher rates of concussion than males in gender-comparable sports, and female sex is a risk factor for post-concussion neuropsychiatric sequelae (Marar et al, 2012; King, 2014). Finally, women are more likely than men to seek medical care and to be exposed to pharmacologic treatment, particularly psychotropic agents that may adversely impact neurologic, cognitive, and sexual function (Whitley and Lindsey, 2009). These latter two examples of sports-related concussions and medication utilization are particularly relevant to the thesis of this review in that they illustrate how behavioral factors under the rubric of gender interact with sex with respect to health outcomes, suggesting that in clinical research the role of both gender and sex are frequently relevant.

The observation that female sex is a risk factor for a wide range of adverse drug reactions bears additional discussion given the readership and mission of this Journal with respect to pharmacology (Parekh et al, 2011; Franconi et al, 2012; Spoletini et al, 2012). Early limitations on recruitment of women to randomized clinical trials and the delay in considering SABV in drug studies have resulted in a number of cautionary tales. Findings from the Physicians’ Health Study that included over 20,000 male physicians studied in the 1980s, indicated that low-dose aspirin was associated

Neuropsychopharmacology
with a reduced risk of myocardial infarction and a slight increased risk of stroke (1989). It was decades later that a follow-up study conducted in 39,000 female participants showed alarmingly disparate findings. In females, low-dose aspirin was associated with a greater risk of stroke with no overall effect on risk for myocardial infarction (Ridker et al., 2005). Although the 1993 NIH policy encouraging the inclusion of women in clinical trials has contributed to a more equal number of males and females in phase III trials, examination of medication efficacy and adverse drug reactions by sex are still lacking, leading to continued evidence of increased risk for women (Mazure and Jones, 2015). Referable to this point, in 2013 the FDA-approved labeling changes for dosing of zolpidem, a drug approved for the short-term treatment of insomnia decades earlier. The recommended starting dose was reduced in women, as metabolism of the drug is slower in females, leading to higher drug concentrations and morning drowsiness in women compared with men. In fact, governmental reports indicate that out of the 10 prescription drugs removed from the US market between 1997 and 2001, 8 were found to have greater adverse effects in women compared with men.

Although our understanding of SABV in many disease states has been hindered by factors from limited financial and physical resources to the belief that sex differences are not important or do not exist (Cahill, 2014), research examining how SABV applies to transgender populations is nonexistent for most medical conditions. Although this topic is of growing and critical importance, the current review focuses primarily on SABV in organisms/individuals for whom the gonadal hormonal milieu is consistent with their chromosomal sex.

Sex Differences Related to Sex Chromosomes and Hormones and the Importance of Timing

Importantly, the mammalian brain develops in the face of combined and opposing forces contributing to resiliency or vulnerability. The sexually dimorphic developing brain is organized in large part by developmental hormone exposure, with males experiencing elevated testosterone levels during the process of normal testes development, largely in utero, although the exact timing of this is species-dependent. Aromatization of this testosterone to estradiol in the brain drives masculinization, an active process affecting cell differentiation and connectivity in the brain in rodents (Figure 4; McCarthy, 1994; McCarthy, 2016). Estrogenic involvement in cell death and cell birth in the developing nervous system is a critical component in programming the sexually dimorphic brain, as is the epigenetic process of DNA methylation in feminization of the female brain (McCarthy et al., 2009; Nugent et al., 2015). Aromatization of testosterone to estradiol occurs in primates, including humans, but does not necessarily have a major role in masculinization of the primate brain. Instead, both testicular production of testosterone and intact androgen receptors are necessary to feminize the human brain as evidenced by the feminine phenotype of XY individuals with complete androgen insensitivity syndrome and masculine phenotype of those with aromatase dysfunction (Imperato-McGinley et al., 1982; Grumbach and Auchus, 1999). Certainly, evidence in humans also suggests a strong correlation between fetal testosterone levels and stereotyped activity with neurodevelopmental disease risk and adult cognitive and behavioral stress reactivity, supporting the importance of the processes involved in establishing the sexually dimorphic
brain (Chura et al, 2010; Lombardo et al, 2012; Lombardo et al, 2012; Baron-Cohen et al, 2015). Programming of important regulatory brain regions, including the neuroendocrine hypothalamus, via such steroid hormone effects on cell migration patterns during early development likely also contributes to sex differences in how the individual responds to environmental perturbations throughout life (Bale et al, 2010).

In rodent models, studies have been able to manipulate this critical organizational window to mechanistically examine its importance in sex differences in long-term stress programming. For instance, adult sex differences in hippocampal expression of the glucocorticoid receptor, critical for the negative feedback and modulation of the HPA stress axis, was disrupted in females masculinized at birth by a single injection of testosterone, supporting the importance of the male testosterone surge in normal wiring of sex differences in stress pathways (Bingham and Viau, 2008; Goel and Bale, 2008; Goel et al, 2011). In addition to this early critical period of programming, the rise in gonadal hormones beginning in puberty interacts with these organizational changes and exerts modulatory actions on neurotransmitter systems critical in regulation of sex differences in stress responsivity. Ultimately then, this is an excellent example of how an important coordination must occur between organizational and activational periods of hormone exposure with additional factors from the sex chromosomes to orchestrate a ‘normal’ stress phenotype for each sex. Therefore, neuropsychiatric disease predisposition may result from a mismatch between these processes due to perturbations during early development (Karatsoreos and McEwen, 2013; Karatsoreos and McEwen, 2013; Pembrey et al, 2014).

Not surprisingly, chromosomal sex also contributes to brain development in a hormone-independent manner. The four core genotype mice in which the testis-determining gene, Sry, has been removed from the Y chromosome and is now autosomal, are an interesting tool with which measured outcomes can differentiate sex chromosome complement from gonadal sex (Quinn et al, 2007; Arnold, 2009). The ‘four core genotypes’ refer to the production of mice that are one of four genotypes—XXsry, XX, XYsry, and XY. Studies using these mice have provided a unique and innovative strategy by which labs with interests from addiction to reproductive behaviors have been able to address important questions regarding sex differences independent of gonadal hormones. For example, Jane Taylor’s group, as discussed below, has used these mice to uncover a critical role of sex chromosomes in producing significant differences in food-reinforced habit formation and goal-directed behaviors dependent on the presence of two X chromosomes, and independent of gonadal hormone status (Quinn et al, 2007; Barker et al, 2010). Thus, from these animal studies, we learned that behaviors relevant to diseases such as addiction and obesity are affected by X-linked genes and can be independent of gonadal hormones. Such findings are critical to consider particularly in the context of hypogonadal states, which characterize the prepubertal and postmenopausal windows, or in the context of exogenous use of reproductive hormones that are inconsistent with the individual’s chromosomal sex.

Studies of individuals born with sex chromosome complements that do not conform to XX or XY have informed our understanding of the individual role of the X and Y chromosome in human brain development. Neuroimaging studies comparing children in early puberty with Turner syndrome (XO) and Klinefelter’s syndrome (XXY) to age-matched controls, indicate a dose-dependency of the X chromosome to decreased parieto-occipital and increased temporo-insular gray matter volumes (Hong et al, 2014; Hong and Reiss, 2014). A similar study also reported smaller hippocampal volumes in XXY boys compared with age-matched controls (Rose et al, 2004). Interestingly, XYY karyotypes that do not present with a detectable phenotype have normal behavior and cognitive abilities, and normal serum testosterone levels, suggesting that, although the dosage of X-linked genes is important in normal development, this may not be the case for Y-linked genes. However, it should be noted that it is inherently difficult to completely dissect the gonadal hormone and sex chromosome effects on the developing brain as both XO and XXY children are also typically deficient in gonadal hormone production. Further, the testis-determining factor, Sry, as discussed above, is a Y chromosome gene and a necessary transcription factor for the bipotential gonad to develop into a testis, a necessity for testosterone production during neonatal and pubertal developmental organizational periods (Capel, 1998).

We know that the timing or ‘critical periods’ of brain development and maturation demarcate points where the plastic brain undergoes marked changes that can be distinctly feminine or masculine. It is within these sexually dimorphic nuclei that cells are increasing or decreasing in size, synapse and glia numbers changing, and dendritic spines sprouting at a rapid rate—all of which point to a robust and dynamic state. Therefore, sex-specific changes
may present an opportunity where the normal trajectory becomes most vulnerable to influences that can be redirected or reprogrammed. For instance, in both humans and rodents, several brain regions involved in stress and corticotropin-releasing factor (CRF)-mediated HPA axis regulation and emotional affect, such as the prefrontal cortex (PFC), amygdala, and hippocampus undergo significant maturation during childhood and adolescence. These areas also demonstrate significant sex differences in both structure and function into adulthood, and may be related to varying rates of maturation between males and females.

Recent findings from the Philadelphia Neurodevelopmental Cohort reveal significant nonlinear effects of sex and age/pubertal development on cerebral perfusion with pre- and peri-adolescent females having lower cerebral blood flow (CBF), but by mid-adolescence (15–17 years old) males were lower, and by late-adolescence (17–22 years old) CBF in females began increasing with age (Satterthwaite et al., 2014). These sex differences in CBF were most pronounced in the default mode network, executive system, and regions critical to affective processing. CBF, a critical property of basic brain physiology, has been shown to vary by sex in older adults, is linked to regional brain metabolism, is altered by cognitive activity and load, and is abnormal in various neuropsychiatric conditions (reviewed by Satterthwaite et al., 2014). Sex differences in such a fundamental aspect of brain physiology emphasize the consideration not only of SABV but also of developmental trajectories when conducting human neuroimaging research. Reproductive hormones have pronounced impact on systemic and central vasculature and thus cerebral perfusion, which is particularly important in research focusing on CNS damage under conditions of ischemia (Sorensen et al., 2001). Considering the later end of the life span when cerebral vascular disease and risk for ischemia are more likely, McCollough, Arnold, and colleagues research using the four core genotype mouse model (described above) indicates that sex differences in ischemic stroke sensitivity is influenced primarily by circulating gonadal hormones, not sex chromosome complement (Manwani et al., 2015). Estradiol supplementation was associated with robust CNS protection in an ischemic stroke model in females, while the effect was muted in males, suggesting that both organizational and activational effects of reproductive hormones should be considered under this context.

Similarly, structural neuroimaging studies in humans emphasize the importance of considering SABV and development when considering the integrity of critical limbic structures such as the amygdala, hippocampus, and PFC. Amygdala volume increases significantly more in adolescent boys, whereas hippocampal volume increases faster in girls, especially within the CA1 subfield (Satterthwaite et al., 2014). Moreover, the thinning of frontal cortical gray matter associated with adolescence occurs earlier in females, which may be related to an earlier onset of puberty. Interestingly, these same limbic and forebrain regions are also extremely sensitive to stress (McEwen and Morrison, 2013). As glucocorticoids are potent modulators of synaptic function and plasticity, it is likely that exposure to stressors during adolescence or impaired CRF pathway function could shape these limbic brain areas critical to emotionality, possibly in a sex-dependent manner given their different rates of maturation (Goel and Bale, 2009).

Unfortunately, little is known regarding the ability of stress across the pubertal transition to alter the developmental trajectory of these brain regions in males and females. In humans, studies have shown the impact of preadolescent stress as being particularly critical for adult right amygdala volumes, with adversity (abuse, neglect, and family stress) occurring at age 11 years, contributing to a 3.5-fold effect above adversity occurring at any other time from birth to 18 years of age (Pechtel et al., 2014). Numerous studies show that alterations in amygdala volume are related to negative emotionality (anxiety, depression, and neuroticism), which characterize psychiatric conditions for which there are prominent sex differences in prevalence, and natural history (Drevets et al., 2008; Mincic, 2015). Unfortunately, sex differences in these neuroimaging outcomes are understudied. As adolescence and young adulthood are periods of increased presentation of mental health disorders, including those that are female-specific such as premenstrual dysphoric disorder and catamenial epilepsy, it will be imperative to better identify and delineate interactions between adolescent brain maturation, environmental influences, and sex to identify points of vulnerability (Guille et al., 2008; Epperson et al., 2012). To do so, neuroimaging studies must have sufficient sample or effect sizes with respect to male and female participants or outcomes, respectively, to address these critical issues in normal and aberrant neurodevelopment. Similar to rodent studies, human subject investigations of stress effects on neurodevelopment would benefit from greater consideration of the context and timing of environmental perturbations.

Knowledge Gained by Including SABV in Studies of Neurodevelopment and Neuropsychiatric Disorders

Translation of many of these preclinical findings to human health and disease has been hindered for obvious reasons. Although not as mechanistically revealing, the Avon Longitudinal Study of Parents and Children (ALSPAC, n = 14 551) initiated in the United Kingdom in 1999 has reported several sex differences with respect to maternal psychosocial distress (anxiety and perceived stress) during pregnancy effects on offspring behavior during infancy and childhood (prepubertal window), and at the age of 18 (Capron et al., 2015). Even when controlling for maternal postpartum affective state and obstetrical factors, maternal anxiety late in gestation was associated most robustly with hyperactivity/inattentiveness in boys and total behavioral/emotional problems in boys and girls assessed at 4 years of age (O’Connor et al., 2003). Later in development, sex differences with respect to maternal antenatal anxiety and cortisol awakening response in a subgroup (n = 1000) of 15-year olds from this cohort were less prominent (O’Donnell et al., 2013). Interestingly, maternal, but not paternal, antenatal depression was associated with the increased risk of anxiety disorders at age 18 years among those individuals who remained in the ALSPAC (n = 4303), suggesting specific effects of prenatal programming, in addition to environmental and genetic factors, for a neuropsychiatric condition that occurs two to three times more frequently in females than males (Capron et al., 2015).
Smaller studies focusing on brain imaging outcomes reveal that maternal report and neuroendocrine evidence of psychosocial distress during pregnancy is associated with increased right amygdala volume among female, but not male, offspring studied at 7 years of age. In this sample, amygdala volume mediated the relationship between levels of maternal cortisol during pregnancy and affective problems during childhood (Buss et al., 2012). Similarly, intergenerational transmission of parental experience of childhood adversity has been reported for both females and males in rodents and for female parents in humans (Moog et al., 2016). Maternal history of childhood trauma is associated with higher levels of placental corticotropin-releasing factor across pregnancy with a steeper rise in levels in the late second trimester to parturition.

Taken together, clinical and basic science research suggests that fetal exposure to environmental insults can alter the trajectory of brain development. Without consideration of offspring and parental sex, our understanding of the intergenerational transmission of life experiences would be greatly limited. We provide additional specific examples of the importance of considering SABV in our discussion of specific neuropsychiatric disorders below. It should also be considered here that it is also the case that even if studies including both males and females find no sex differences in their outcomes, sex differences can exist at many levels. Certainly, one can get to the same destination via different routes. For instance, in neuropsychiatric diseases such as autism, schizophrenia, or depression, although men and women may be given a similar diagnosis, there exist significant sex differences in overall rates, timing of onset, symptom presentation, and treatment efficacy (Heim and Nemeroff, 1999; Heim and Nemeroff, 2001; Sanchez et al., 2001; Goldstein et al., 2002; Heim et al., 2004; Bale, 2006; Brown and Susser, 2008; Bale, 2009; Brown et al., 2009; Heim et al., 2009; Bale et al., 2010; Heim et al., 2010; Brown, 2012; Davis and Pfaff, 2014; Goldstein et al., 2014). Further, mechanistic animal studies modeling endophenotypes of these disorders have demonstrated robust sex differences in the timing of susceptibility to insults to the developing brain where males appear more vulnerable prenatally and females postnatally (Mueller and Bale, 2006; Mueller and Bale, 2007; Ivy et al., 2008; Kapoor et al., 2008; Mueller and Bale, 2008; Kapoor et al., 2009; Ivy et al., 2010; Hsiao and Patterson, 2012). Therefore, although a male and a female may both be diagnosed with the same disease, for instance autism, the means by which that disease was programmed and presented may be sex-specific.

**Male-biased disorders.** Recent studies have identified gender-specific patterns and connectivity, suggesting that the brains of boys and girls begin to develop very differently at an early age, where the brains of boys are structured to increase the connectivity between perception and coordinated action, and that of girls to assist in connectivity between analytical and intuitive processing (Satterthwaite et al., 2014). These differences clearly begin during childhood and are further increased by the hormones and maturational processes occurring during adolescence, and as mentioned previously, modified with during periods of dynamic hormonal change such as those occurring during the menstrual cycle, childbearing, and the perimenopause in women.

Certainly, numerous neurodevelopmental disorders exhibit strong sex biases, including autism spectrum disorder with an overall sex ratio of 4:1 for boys:girls (as reviewed in Newschaffer et al. (2007), Davis and Pfaff (2014), and Gore et al. (2014)) as well as attention deficit hyperactivity disorder with a male to female sex ratio of 3.2:1 (as reviewed in Erskine et al. (2013)). The age of onset of these disorders and their male preponderance, suggests that intra-uterine factors are involved in their pathophysiology. Although the onset of schizophrenia peaks in both males and females after puberty, males are more likely to be affected, to have earlier onset and a more protracted course (Bale and Epperson, 2015). A male, but not female, fetus exposed during the second trimester of gestation to the stress of the 1940 invasion of The Netherlands had increased risk of schizophrenia later in life (van Os and Selten, 1998). This relationship between prenatal exposure to maternal psychosocial stress and risk for schizophrenia in male, but not female, offspring has been replicated in another study, but with first trimester exposure (Khashan et al., 2008). Although there are likely multiple factors contributing to sex differences in disease, sex-specific responses to fetal antecedents occurring during gestational sensitive windows may promote differences in programming trajectories that underlie such disease biases.

Along these lines, functional magnetic resonance imaging studies in men and women diagnosed with schizophrenia confirmed that the expected sex differences in regional volumes were disrupted, including the ratio of orbitofrontal cortex:amygdala, where male schizophrenia patients have a phenotypically more female pattern in these brain regions (Cowell et al., 1996; Goldstein et al., 2002; Gur et al., 2004). Hence, without consideration of outcomes by sex, vital information regarding the origins of neuropsychiatric conditions would be missed. Moreover, identification of those at greatest risk of adverse consequences of maternal psychosocial stress may improve outcomes.

**Female-biased disorders.** To our knowledge, there are few studies examining the impact of maternal psychosocial stress during discrete gestational windows on risk for affective disorders, which occur more frequently in females. Perhaps, this is due to the challenge of following cohorts from gestation to post-puberty when the sex bias for affective disorders in females is revealed. The ALSPAC study, previously mentioned, did not address potential sex differences regarding gestational exposure to maternal anxiety and depression on risk for affective disorders at age 18 years (Capron et al., 2015), nor did an Australian longitudinal cohort study (n = 3,099) examine outcomes by sex of the offspring assessed for internalizing and externalizing behaviors at age 21 years (Betts et al., 2015). Given the expense and time associated with prospective cohort studies, we would suggest mining existing databases, which may have sufficient sample sizes to test targeted hypotheses about potential sex or gender differences on risk for affective disorders and other conditions that occur distal to in utero life. Future prospective studies should consider sex and gender in their study design to be compliant with new NIH policies and to reduce missed scientific opportunities.
The Myth of Increased Female Variability

Now that we appreciate the importance and value of including SABV in research studies, questions frequently arise as to how best to design studies that will detect sex differences. We have included several out-takes (Box 1; Figures 1 and 2) to provide an overview of the estrous and menstrual cycles, and how an investigator can focus on or control for particular hormonal and reproductive states in rodents and humans. Our goal is to eradicate the notion and unnecessary roadblock that has hindered many investigators, namely, that females are ‘messy and add variability to results due to their hormonal cycle’. Although there is an entire field devoted to the importance of the female estrous (rodents) or menstrual (primates) cycle on outcomes from dendritic spine dynamics to cognitive processing, most neurobiological measures where an actual sex difference is found are not driven by changes in female cyclicity, the best evidence for which comes from rodent models that we will focus on first.

Yes, estrogen and progesterone significantly change over the estrous and menstrual cycles (Box 1; Figure 1) and across the reproductive life span (Figure 2). And, yes, gonadal hormones can promote significant changes in important neurobiological measures when specifically examined. However, a comprehensive meta-analysis recently completed by Zucker et al examining the variability within sex in 293 publications that included male and female mice showed that, in fact, females were equally or even less variable in all reported molecular, physiological, and behavioral measures examined than were males (Prendergast et al, 2014). This variability was similar between males and females whether animals were single or group housed. This variability comparison has now similarly been completed in rats as well (Becker et al, 2016). In addition, many behavioral studies have similarly reported that when estrous cycle stage was included as a variable, no significant differences in outcome measures were found, including for the Barnes maze, tail suspension test, forced swim test, elevated plus maze, light–dark box, and open-field tests (Mueller and Bale, 2007; Goel and Bale, 2008; Goel and Bale, 2010; Goel et al, 2011). This is not to say that gonadal hormones are not important in behavioral responses, but rather that the variability within females when comparing their outcomes to males does not explain sex differences.

Therefore, although the NIH and other funding agencies in the United States, Canada, and the European Union have now pushed forth the mandate to include SABV in all grant proposal submissions, there remains confusion as to how to best design and interpret results in which males and females are both included. First, as discussed in the Circumspective by McCarthy and Joel in this issue, to include SABV does not suggest that the researchers are asking nor are they powered to ask, if sex differences are present (Joel and McCarthy, 2016). Including both sexes and examining outcomes for potential differences between sexes in the results should provide insight as to whether major differences would be present and if further studies are warranted. The NIH has made it very clear that they are not requiring that everyone study sex differences nor are they in a position to fund the expenses for doubling of experimental numbers. But including female comparisons in what has been for most fields largely male-centric can only be seen as a benefit, as we consider additional modes and mechanisms in disease risk and resilience across the life span.

So, how then does an experimenter design their studies so as to be inclusive of SABV without doubling their experimental numbers? To start, knowing the literature as to reported sex differences can be very helpful in how to best go about including both sexes in your study or the appropriate justification for focusing on one sex only. Similarly, if you are expecting sex differences, knowing what type of sex difference can make a big difference in your experimental numbers and planned comparisons. For instance, if you are running a Barnes maze behavioral test and knowing that males and females use different strategies to locate the hidden escape box, then you know that the latency for both groups may be different, but the interpretation of the difference is not that females have decreased spatial ability, but rather that their less risk-aversive serial search strategy requires additional travel time (Mueller and Bale, 2007). This is an example of a sexually dimorphic outcome where the measure exists in two forms, one more prevalent in one sex vs the other (ie, females using a serial search strategy, males using a spatial navigation strategy in this test; Joel and McCarthy, 2016). Knowing this, you might design your study to ask about differences within sex for a given manipulation rather than between sexes. As another example, in post-puberty rodents, the HPA stress axis has a significant sex difference in the amount of ACTH and corticosterone produced in response to a given acute stressor (Handa et al, 1994; Handa et al, 1994; Romeo et al, 2004; Romeo and McEwen, 2006; Goel and Bale, 2007; Goel and Bale, 2008; Goel and Bale, 2009; Handa et al, 2009; Weiser and Handa, 2009). This would be an example of a sex difference that is along a continuum, with females producing the same hormones and in the same direction as males (increased following stress), but just at higher levels (Joel and McCarthy, 2016). In this example, one should be aware that when designing studies comparing sexes, most biological responses have floor and ceiling effects. So, although one sex can increase (males) or decrease (females) their stress response more than the other following a given manipulation, this may not constitute an actual sex difference due to the limitations of different starting points for each sex. Last, if you detect high variability specifically in your female groups and want to determine the potential impact of cycle stage, we have included general information for a ‘how to’ in Box 1. Most rodents will move through four stages of the estrous cycle as follows: metestrus, diestrus, proestrus, and estrus. Rats are more predictable in this 4-day cycle than mice tend to be, although this can be strain-, age-, and environment-specific (Caligioni, 2009). Females co-housed together do not necessarily cycle together, but in mice, this group housing can induce a ‘persistent estrus’ state. Estradiol and progesterone levels are highest leading up to proestrus, which promotes reproductive behaviors and ovulation, and then remain relatively low across the remainder of the estrous cycle.

Similarly in humans, consideration of the menstrual cycle (Figure 1) and reproductive status (Figure 2) of the individual may reduce overall noise in the data and enhance the capacity to detect the outcomes between groups. As an example, pre-pulse inhibition (PPI) of the acoustic startle response, a measure of sensorimotor gating, is diminished...
among individuals with schizophrenia as well as in healthy women during late pregnancy and the luteal phase of the menstrual cycle (Braff et al, 2001). PPI response has historically been considered a potential endophenotype for schizophrenia, and a number of drug development studies have focused on the capacity of a given drug to modulate the PPI as a model for potential efficacy in the treatment of schizophrenia (Braff and Light, 2005; Javitt et al, 2008). Hence, investigations of PPI in the pathophysiology and treatment of schizophrenia should take into consideration the reproductive status of the individual as gonadal steroids can modify PPI in both men and women. This is but one example of an area of neuropsychiatric research for which an investigator should consider SABV or they risk obtaining a finding that is specific to reproductive aged males only (before the age-related decline in androgens; Figure 2).

How then do clinical and translational investigators know when and how to consider SABV and design their investigation in a manner to directly test for sex differences? We would argue that all research focusing on normal physiology and disease states or conditions which occur in both sexes should be powered to detect a potential sex difference in outcomes of interest or at the very least sex/gender should be included as a covariate. If both sexes will not be included in the investigation in question, the onus is upon the investigator to justify the exclusion of one sex or the other. This will be relatively simple when studying conditions such as breast cancer, which occur in both sexes, but 99% of cases are women, or pregnancy states for obvious reasons. In contrast, coronary heart disease (CHD) is roughly equal between men and women between the ages of 40 and 59 years (Mozaffarian et al, 2015). Therefore, it may be difficult to justify a male only study of CHD among younger adults. Further, estradiol has the protective effects against CHD among healthy young female adults, suggesting that reproductive status, as well as age, is crucial to consider when investigating CHD.

Figure 1 illustrates a typical menstrual cycle with respect to length and reproductive hormone levels among premenopausal women. If there is a reason to consider the potential effects of gonadal hormones on an outcome of interest such as PPI, it is most straightforward to study all female participants in the early to mid-follicular phase when the vast majority of women will have relatively low levels of estradiol and progesterone as well as the progesterone neurosteroid metabolite, allopregnanolone (ALLO). By day 8 of the menstrual cycle (with day 1 anchored to the first day of menstruation), estradiol begins to rise but does not peak until ovulation. Finally, the luteal phase of the menstrual cycle is characterized by a more complex hormonal milieu that is less predictable between cycles and between women. Young adolescent girls and women with disorders such as polycystic ovarian syndrome frequently have anovulatory cycles, characterized by low progesterone levels. Such cycles disrupt the normal hormonal milieu of the luteal phase as well as the length of the cycle, making timing of an investigation or procedure with respect to the desired phase more challenging.

For investigators who wish to isolate the impact of estradiol on outcomes of interest in naturally cycling women, test days should occur during the first week of the onset of menstruation and between days 12 and 15 of an idealized 28-day cycle. In this case, within subject design is ideal. However, some investigators such as Mohammed Milad and his laboratory have been successful in examining sex, reproductive hormones, and oral contraceptive pill (OCP) effects on their outcomes of interest (primarily fear conditioning and extinction) by studying men and women without regard for the hormonal status of the female on a given test day (Hwang et al, 2015). They parsed the women into groups based upon blood hormone levels or self-report of timing of last menstrual period and report of OCP use. This procedure is most effective when there are few constraints on access to study equipment or facilities, but may be less effective in costly neuroimaging studies if a specific hormonal milieu is the goal. The gold standard for defining menstrual cycle phase or effects of specific levels of hormones on outcomes is to obtain blood samples for hormonal staging on each test day.

For those who are interested in capturing the luteal phase of the cycle, it is best to require participants to track their menstrual pattern for at least one cycle and to determine the timing of ovulation using urine luteinizing hormone (LH) kits. These kits can be somewhat difficult to interpret, and women should follow the instructions for a specific kit closely. In our experience, women frequently question whether a LH detection line is dark enough to be positive and we now ask all women to take a cell phone picture of their results to forward to our investigators for confirmation.

CONCLUSIONS

In summary, the NIH policy to consider SABV is a mandate that impacts every investigator who plans to seek federal funding for their research. However, we contend that the most important reason to consider sex and gender in biomedical research is that it improves the quality of the science, and the safety and efficacy of treatments for human disorders. In addition, there is great potential to gain mechanistic insight from studies where sex differences are identified. The antiquated myth that females are ‘messy’ and make research more difficult is not supported by the evidence, and articles such as this one, and those from other investigators who have been considering sex and gender in their research for decades, will help promote the inclusion and importance of SABV in biomedical research (Bale and Epperson, 2015; Miller et al, 2015; Becker and Koob, 2016; Joel and McCarthy, 2016).

FUNDING AND DISCLOSURE

The work discussed in this review was funded in part by grants from the NIH MH073030 (Bale), MH091258 (Bale), MH087597 (Bale), MH104184 (Bale), MH108286 (Bale), MH099910 (Epperson and Bale), AG048839 (Epperson), and DA030301 (Epperson). Dr Epperson acknowledges the receipt of research grant support from Shire and Sage Therapeutics, stock investments in Eli Lilly, Pfizer, Abbott, Abbvie, Merck, and Johnson and 20 Johnson, as well as consulting to Sage Therapeutics, Forest Laboratories, and Asarina Pharmaceuticals. The authors declare no conflict of interest.
REFERENCES


Sex as a biological variable
TL Bale and CN Epperson


Joel D, McCarthy MM (2016). Incorporating sex as a biological variable in neuropsychiatric research: where are we now and where should we be? Neuropsychopharmacology (e-pub ahead of print; doi:10.1038/npp.2016.79).


Internalizing and externalizing behaviors predict elevated inflammatory markers in childhood

Natalie Slopen a,b, Laura D. Kubzansky b, Karestan C. Koenen c,*

a Center on the Developing Child, Harvard University, Cambridge, MA, USA
b Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA
c Department of Epidemiology, Columbia University, Mailman School of Public Health, New York, NY, USA

Received 6 May 2013; received in revised form 10 July 2013; accepted 22 July 2013

Summary

Background: Children with behavior problems, such as internalizing or externalizing disorders, are at increased risk for poorer physical health in adulthood. Inflammation has been posited as a potential biological mediator underlying this association. However, it is unclear how early in development associations between behavior problems and inflammation may be detected, and whether associations are present for both internalizing and externalizing behaviors in prepubertal children.

Methods: Using data from children in the Avon Longitudinal Study of Parents and Children, we examined associations between behavior problems at age 8 (assessed via the parent-report Strengths and Difficulties Questionnaire) and inflammatory markers assessed at age 10. Inflammatory markers included C-reactive protein (CRP; n = 4069) and interleukin-6 (IL-6; n = 4061). We further evaluated whether body mass index (BMI) mediated associations, and tested for potential reverse causality by considering whether age 10 inflammation was associated with changes from initial levels to age 12 behavior problems.

Results: After adjusting for relevant covariates, age 8 externalizing behaviors were associated with elevated CRP at age 10, and age 8 internalizing and externalizing behaviors were associated with elevated IL-6 at age 10 (p’s < 0.05). We found no evidence that observed associations were mediated by BMI or that inflammatory markers at age 10 were associated with increased internalizing or externalizing behavior problems at age 12.

Conclusions: These findings document an association between behavior problems and elevated concentrations of CRP and IL-6 at 10 years. Heightened inflammation in childhood may be a pathway through which early behavior problems increase risk for adult chronic diseases.

* Corresponding author at: Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 720G, New York, NY 10032, USA. Tel.: +1 212 305 7482; fax: +1 646 317 0013.
E-mail address: kck5@columbia.edu (K.C. Koenen).

0306-4530/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.psyneuen.2013.07.012
The childhood origins of adult chronic disease are well-established (Garner et al., 2012), with childhood behavior problems, such as internalizing and externalizing disorders, emerging as potential risk factors for poor adult physical health (Odgers et al., 2007; Jokela et al., 2009; Appleton et al., 2011). However, we have a limited understanding of how early in life childhood behavior problems may become linked to pathophysiological mechanisms that produce excess risk for poor adult physical health. Both internalizing and externalizing behaviors are associated with dysregulated hypothalamic–pituitary–adrenal axis (HPA) (Cappadocia et al., 2009; Lopez-Duran et al., 2009) and sympathetic nervous system (SNS) (Matthys et al., 2013; Obradovic and Boyce, 2012) function, which could lead to the initiation and/or sustained release of pro-inflammatory cytokines (Miller and Cole, 2012). The resulting inflammation may serve as one biological mechanism by which childhood behavior problems result in excess disease risk decades later. The aim of this study was to investigate whether childhood internalizing and externalizing behaviors predict elevated chronic inflammation two years later, as indicated by C-reactive protein (CRP) and interleukin-6 (IL-6) levels, and focused on whether such effects may be evident even in pre-pubertal children.

CRP and IL-6 are well-established markers of systemic inflammation (Danesh et al., 2004; 2008). CRP is the principal downstream mediator of the acute phase response to inflammatory processes occurring in the body (e.g., a rise in IL-6). Prospective studies show that elevated CRP and IL-6 are associated with greater atherosclerosis and incident coronary events in adults (Danesh et al., 2008; Kapothe et al., 2010). Thus, elevated levels of these markers can provide an early indicator of disease risk even if causal associations with disease outcomes are unclear (C Reactive Protein Coronary Heart Disease Genetics Collaboration, 2011). Prior research supports a relationship between childhood CRP and IL-6 levels with other well-established cardiovascular and atherosclerotic risk factors in children (Cook et al., 2000; Jarvisalo et al., 2002); and, elevated CRP in childhood tracks into adulthood (Juonala et al., 2005).

Internalizing behavior problems refer to symptoms of anxiety, depression, or withdrawal, while externalizing behavior problems refer to symptoms of hyperactivity, conduct problems, or aggression (Goodman et al., 2010). Several epidemiologic studies suggest a long-term influence of childhood internalizing and externalizing behaviors on inflammation in adulthood (Odgers et al., 2007; Appleton et al., 2011, 2012); however, few studies have examined these behavior problems in relation to inflammation among children prior to puberty (Bartlett et al., 1995; Brambilla et al., 2004; Caserta et al., 2011). Notably, all previous studies using pre-pubescent samples examined major depression or depressive symptoms, but not other forms of internalizing or externalizing behaviors. Most studies using adolescent samples also focused on depression (Targum et al., 1990; Shain et al., 1991; Birmaher et al., 1994; Schleifer et al., 2002; Gabby et al., 2009; Copeland et al., 2012; Henje Blom et al., 2012; Miller and Cole, 2012). Evidence on the relation between depression and immune parameters in children or adolescents is inconsistent. Some studies show no differences in immune parameters by depressive status (Targum et al., 1990; Shain et al., 1991; Birmaher et al., 1994), while others indicate at least modest evidence of immune-related abnormalities associated with depressive symptoms (Bartlett et al., 1995; Schleifer et al., 2002; Brambilla et al., 2004; Gabby et al., 2009; Caserta et al., 2011; Copeland et al., 2012; Henje Blom et al., 2012; Miller and Cole, 2012). Only a few studies have examined externalizing behavior problems in relation to immune parameters among adolescents (Targum et al., 1990; Birmaher et al., 1994; Pajer et al., 2002). Similar to studies examining depression, findings for externalizing behavior are inconsistent. Some evidence suggests no association between externalizing behavior and immunological abnormalities (Targum et al., 1990; Birmaher et al., 1994), while some evidence supports a positive association (Pajer et al., 2002).

Understanding the association between behavior problems and inflammation in children is constrained by six limitations in extant research. First, most prior studies used small convenience samples. Selection bias and limited power may contribute to discrepant results. Second, few existing studies have compared effects of a range of both internalizing and externalizing behaviors on inflammation. Thus, it is unknown whether adverse effects of early behavior problems on health are general or specific. Third, across the prior literature on behavior problems or depression in youth and immune parameters, most studies examining multiple outcomes have reported differing associations across immune markers. This makes it challenging to interpret overall findings, and to draw conclusions about the reliability of reported associations. Fourth, few existing studies examined potential differences by sex, even though some evidence suggests this may be important among preadolescents (Caserta et al., 2011). Fifth, body mass index (BMI) has not been examined as a mediator of observed associations. Childhood internalizing and externalizing behaviors are associated with elevated body mass in adolescents (Goodman and Whitaker, 2002; Cortese et al., 2008), and BMI is associated with inflammation (Lambert et al., 2004). Although some studies in youth have included BMI as a covariate, BMI has rarely been examined as a pathway. Finally, most prior studies collected information about behavior and inflammation at a single time-point and thus could not explore directionality. Studies of adults have detected a bi-directional association between inflammation and depression (Howren et al., 2009), however it is less clear whether a bidirectional association occurs in younger populations (Copeland et al., 2012; Miller and Cole, 2012).

In the present study, we capitalized on data from a large population-based cohort to test the hypothesis that higher internalizing and externalizing behaviors are associated with higher levels of CRP and IL-6 among children prior to adolescence. We examined behavioral problems at age 8 in relation to inflammatory outcomes at age 10 to evaluate these associations prior to puberty, and to ensure our predictor variable was measured temporally prior, but in close proximity, to our main outcome measure. We also hypothesized that positive associations would be mediated by elevated BMI, for reasons described above. Finally, as an exploratory analysis, we tested the possibility of reverse causality by evaluating whether inflammation at age 10 could predict increased internalizing or externalizing behaviors among children at age 12, taking account of initial levels of behavior problems. Consistent with prior research, we included relevant covariates in our models to reduce potential confounding, including
sex, ethnicity, age, current medication use, maternal education at gestation, and family income at age 8.

1. Materials and methods

We performed a secondary analysis of existing data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a population-based longitudinal study of children born to mothers who resided in Avon County, England while they were pregnant (Golding and the ALSPAC Study Team, 2004; Boyd et al., 2012). ALSPAC was initiated to study mechanisms by which social, biological, and environmental factors influence pregnancy outcomes and child mental and physical health. 14,541 pregnant women with estimated delivery dates between April 1991 and December 1992 agreed to participate; this included approximately 85% of eligible participants. A total of 13,988 children were still alive at 12 months, and the study aimed to follow these children. Questionnaires were periodically mailed to parents during the pregnancy period and beyond; beginning at age 7, the sample was invited for bi-annual clinic visits. The ALSPAC Law and Ethics Committee and Local Research Ethics Committee approved the study. In-depth methodological details are described on the study website (www.alspac.bris.ac.uk) and have been reported in other publications (Golding and the ALSPAC Study Team, 2004; Boyd et al., 2012). Children who dropped out of the sample were more likely to have behavior disorders and to have mothers with more financial difficulties and lower education (Wolke et al., 2009).

Surviving children whose parents provided consent for continuing participation and with valid contact information were recruited for biological assessments at the Age 9 Clinic (mean age at visit = 9.8 years; hereafter we refer to the outcomes as an age 10 measure). A total of 7725 children attended the clinic (62% of invited participants), and 5086 of these children were willing to participate in the venipuncture protocol. Active infection can cause elevations in inflammation unrelated to behavior problems, therefore we excluded individuals who reported infection in the 7 days prior to blood collection (431 children). Children with a valid measure of CRP or IL-6 were eligible for this analysis (N = 4655 and 4647, respectively). The sample for the main analyses included 4069 children who met the following inclusion criteria: data on behavior problems at age 8, data on all relevant covariates, and a valid assessment of CRP or IL-6 at age 10 (models of CRP, n = 4069; models of IL-6, n = 4061).

In a sub-analysis, we examined if inflammation at age 10 was associated with increased behavior problems at age 12. This sub-analysis included 3634 individuals with assessments of CRP or IL-6 at age 10, complete data on control covariates, and data on behavior problems for the age 8 and age 12 measures; models of CRP, n = 3634; models of IL-6, n = 3627). In Appendix 1, we summarize the timing of assessment for the exposures and outcomes in our analyses.

1.1. Measures

1.1.1. Inflammatory outcomes

Blood samples were collected using standard methods. Concentration of CRP (mg/L) was measured by automated particle-enhanced immunoturbidimetric assay (Roche, UK), and concentration of high sensitivity IL-6 (pg/ml) was measured by enzyme-linked immunosorbent assay (R&D systems, Abingdon, UK). Assay coefficients of variation for both outcomes were <5%. Both CRP and IL-6 were log-transformed for regression analyses. Following standard practices, observations greater than three standard deviations above or below the sample mean were winsorized to prevent bias due to outliers in linear models (IL-6, N = 44; CRP, N = 39) (Wilcox and Keselman, 2003).

1.1.2. Behavior problems

Behavior problems were assessed via mail questionnaire using the parent-rated Strengths and Difficulties Questionnaire (SDQ) (Goodman, 2001), a validated assessment of behavioral symptomatology. Mothers were asked to rate 25 items describing their children's behavior (in the past 6 months) as “not true”, “somewhat true”, or “certainly true”. The SDQ includes one subscale to assess prosocial behavior, and four subscales to assess problematic emotions and behaviors, including conduct, hyperactivity, peer, and emotional problems (Goodman, 2001). The reliability of the SDQ is considered satisfactory (Goodman, 2001), according to measures of internal consistency (mean Chronbach’s alpha across subscales = 0.73), cross-informant correlation (mean across subscales = 0.34) and test–retest stability after 4–6 months (mean across subscales = 0.62). For analyses in epidemiological samples within the general population, there is theoretical and empirical evidence to support two broad categorizations of problems, an “internalizing” subscale which combines the emotional and peer problems subscales (10 items; e.g., often seems worried; often unhappy, depressed or tearful; rather solitary, prefers to play alone), and an “externalizing” subscale which combines the conduct and hyperactivity problems subscales (10 items; e.g., restlessness; often loses temper; often lies or cheats) (Goodman et al., 2010); therefore, we used the internalizing and externalizing subscales in our main analyses (possible range for both scales: 0–20). The full scale is available at http://www.sdqinfo.com. The internalizing and externalizing scales were transformed into z-scores for analysis to improve interpretability.

Primary analyses utilized SDQ symptoms reported in the Age 8 questionnaire. For participants missing SDQ data from the Age 8 questionnaire, we created scores from identical questions in the Age 7 questionnaire (n = 392, 9.6%). The overall mean age at the time of report was 8.1 years. In additional analyses, SDQ scores from the Age 11 questionnaire were used in models to examine whether inflammation was associated with subsequent increases in internalizing or externalizing behaviors. The mean age at the Age 11 questionnaire completion was 11.7 years, thus we refer to these assessments as age 12 behavior problems (see Appendix 1).

1.1.3. BMI

At the Age 9 clinic (mean age of respondent = 9.81 years, hereafter referred to as age 10 BMI), staff measured height using the Harpenden Stadiometer to the last complete millimeter, and unclothed weight to 0.1 kg. Child BMI was calculated as weight (kg)/length (m)^2. BMI was transformed to BMI z-scores following UK 1990 BMI population reference data (Cole et al., 1995).
1.1.4. Covariates

We included a number of covariates known to be associated with variation in inflammation (O’Connor et al., 2009) or that have been used as covariates other studies of inflammation in youth (Caserta et al., 2008; Copeland et al., 2012; Fuligni et al., 2009b; Murasko, 2008; Slopen et al., 2012). During pregnancy, mothers reported their highest level of education (below O-level, O-level only, A-level only, university degree or more, missing). After the child was born, mother’s reported the child’s sex (male, female), and ethnicity (white, non-white; based on mother’s report for herself and her partner). In the Age 8 questionnaire, mothers reported the weekly household income (<£100, £100–£199, £200–£299, £300–£399, >£400, unknown). And, at the time of the blood draw, children reported their age (recorded in months) and parents reported if the child was taking any medication (yes/no).

1.2. Analyses

First, we used linear regression models to examine associations of internalizing and externalizing behaviors assessed at 8 years with CRP and IL-6 assessed at age 10. To show any change in the main associations of interest after standard control covariates were introduced, we used a two-step model building approach. We started with a simple model that included controls for sex and age at blood collection. We then fit models additionally adjusted for ethnicity, current medication use, maternal education at gestation, and family income at age 8. Second, we considered potential sex differences in associations by testing interaction terms for sex by internalizing or externalizing behaviors. Third, we used a bootstrapping approach to examine the extent to which observed associations between behavior problems and inflammation were mediated by BMI (Preacher and Hayes, 2008). This method provided an estimate of the indirect effect of BMI, which describes the proportion of the association between behavior problems and inflammation that is explained by BMI (i.e., the degree of mediation that occurred) (Preacher and Hayes, 2008). See Appendix 2 for a diagram and further explanation of the mediation analyses. If the confidence interval for an indirect effect does not cross zero, this indicates that significant mediation occurred.

Finally, to examine if inflammation was associated with an increase in later behavior problems, we estimated linear regression models with CRP and IL-6 at age 10 as the independent variables, and behavior problems at age 12 as the dependent variables (in separate models). These models adjusted for internalizing or externalizing behaviors at age 8, in addition to the same covariates used in the primary analyses. By adjusting for baseline problems, we obtained an estimate of the change in behavior from initial assessment to age 12 assessment associated with age 10 levels of CRP and IL-6. All analyses were conducted in SAS v.9.2, and statistical significance was established at p < 0.05.

2. Results

Table 1 presents a description of the sample. Approximately half of the sample was female, and the mean age was 9.8 years at the time of blood collection. Consistent with the region, 95% of the sample was White. There was heterogeneity by socioeconomic status. At the time of gestation, 19% of children had mothers who had not completed an O-level education; at 8 years, 46% of children resided in households that earned £400 or more per week.

In initial analyses we did not find evidence for significant differences by sex. Thus, in all further analyses, we present results combined across girls and boys.

2.1. Associations between behavior problems and CRP and IL-6

Table 2 presents parameter estimates for the associations between internalizing and externalizing behavior at age 8.
and CRP and IL-6 at age 10. The pattern of significant associations for minimally adjusted models and fully adjusted models are similar, although the coefficients are slightly attenuated in the fully adjusted models. Considering fully-adjusted models with log-CRP as the outcome, externalizing behavior was associated with elevated CRP ($\beta = 0.04$, $p = 0.04$, model $R^2 = 0.05$), while the association between internalizing behavior and CRP was marginally statistically significant ($\beta = 0.03$, $p = 0.06$, model $R^2 = 0.05$). In the fully adjusted models with log IL-6 as the outcome, both internalizing and externalizing behaviors were associated with elevated levels of IL-6 (i.e., internalizing: $\beta = 0.05$, $p < 0.01$, model $R^2 = 0.03$; externalizing: $\beta = 0.04$, $p \leq 0.01$, model $R^2 = 0.03$).

### 2.2. BMI as a potential mechanism

For the significant associations described above, we evaluated BMI z-score as a potential mediator. Age 8 externalizing behavior was associated with BMI z-score ($\beta = 0.04$, $p = 0.04$), but there was no association for age 8 internalizing behavior ($\beta = 0.01$, $p = 0.76$). BMI z-score was positively associated with CRP ($\beta = 0.39$, $p < 0.0001$) and IL-6 ($\beta = 0.15$, $p < 0.0001$). However, the indirect effect (i.e., the proportion of the association that can be explained by BMI) in each of the mediation models was not significant, thereby failing to provide evidence that the observed associations were mediated by BMI (for table of coefficients, see Appendix 3).

### 2.3. Associations between CRP and IL-6 and later behavior problems

Table 3 presents associations between CRP and IL-6 at age 10 and change in internalizing and externalizing behaviors between the earlier assessment and the age 12 assessment. Neither CRP nor IL-6 was associated with change in internalizing or externalizing behaviors over time.

### 3. Discussion

In a large cohort of pre-adolescent children, we found behavior problems were associated with elevation in inflammatory markers, and associations were maintained after taking account of a range of relevant covariates. To our knowledge, this is the first large population-based study to examine internalizing and externalizing behaviors together in relation to inflammation in a pre-adolescent sample. Elevated CRP levels were strongly associated with externalizing behavior and more weakly associated with internalizing behavior, while both internalizing and externalizing behaviors were associated with elevated IL-6 levels. We found no evidence for differential associations by sex, or that observed associations were mediated by BMI. While a true prospective examination was not possible in the primary

### Table 2  Associations between age 8 internalizing and externalizing behaviors and C-reactive protein and Interleukin-6 levels at age 10.

<table>
<thead>
<tr>
<th>Outcome: log C-reactive protein</th>
<th>Simple model$^a$</th>
<th>Adjusted for covariates$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internalizing z-score</td>
<td>$\beta$ (SE)</td>
<td>$p$-Value</td>
</tr>
<tr>
<td>0.04 (0.02)</td>
<td>0.02</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>Externalizing z-score</td>
<td>0.05 (0.02)</td>
<td>0.01</td>
</tr>
<tr>
<td>Outcome: log interleukin-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internalizing z-score</td>
<td>0.06 (0.01)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Externalizing z-score</td>
<td>0.05 (0.01)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

$^a$ Simple linear regression models are adjusted for age and sex.

$^b$ Covariates include sex, ethnicity, age, current medication use, maternal education at gestation, family income at age 8. Sample sizes vary slightly due to missing information on inflammatory markers ($n = 4069$ for CRP and $n = 4061$ for IL-6). SE, standard error.

### Table 3  Associations between age 10 C-reactive protein and interleukin-6 and change in internalizing and externalizing behaviors between initial assessment and 12 years.$^a$

<table>
<thead>
<tr>
<th>Outcome: internalizing behavior (z-score)</th>
<th>$\beta$ (SE)</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log C-reactive protein</td>
<td>0.00 (0.01)</td>
<td>0.81</td>
</tr>
<tr>
<td>Log interleukin-6</td>
<td>0.01 (0.02)</td>
<td>0.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome: externalizing behavior (z-score)</th>
<th>$\beta$ (SE)</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log C-reactive protein</td>
<td>-0.02 (0.01)</td>
<td>0.13</td>
</tr>
<tr>
<td>Log interleukin-6</td>
<td>-0.02 (0.02)</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Sample sizes vary slightly due to missing information on inflammatory markers ($n = 3634$ for CRP and $n = 3627$ for IL-6). BMI, body mass index; SE, standard error.

$^a$ Models adjusted for age, sex, ethnicity, current medication use, maternal education at gestation, family income at age 8, age 10 BMI z-score, and age 8 internalizing/externalizing behavior.
analysis due to lack of available information on initial levels of inflammation, the finding that elevated inflammation was not associated with subsequent increases in internalizing or externalizing behaviors over time reduces concerns about reverse causality. Taken together, our findings suggest that inflammation may be a pathway through which childhood behavior problems increase later risk for chronic diseases in adulthood.

Our findings extend a small set of studies that have examined depression or other behavior problems in relation to CRP or IL-6 in pre-adolescent or adolescent samples (Gabbay et al., 2009; Caserta et al., 2011; Copeland et al., 2012; Henje Blom et al., 2012; Miller and Cole, 2012). Two recent studies have found associations between depression and elevated CRP among adolescents (Copeland et al., 2012; Miller and Cole, 2012); while the present study demonstrated an association between internalizing behavior and elevated CRP (after adjustment for family socioeconomic status) that was somewhat weaker than expected, it is consistent with findings from these studies. Our results build on the prior literature by also examining externalizing behavior in relation to CRP and IL-6 in pre-adolescent children, and documenting a strong association. Our findings further support a uni-directional relationship between behavior problems and later inflammation, similar to results from a prospective cohort study that tested bi-directional longitudinal associations between CRP and depression in a slightly older sample over a longer period of time (Copeland et al., 2012). The alternative pathway (i.e., inflammation to later depression) has been supported in older age groups (Gimeno et al., 2009); it is possible that this pathway is slower to develop or simply less common so that preadolescents and adolescents do not show the same pattern reported among adults.

Studies on behavior problems in relation to inflammation in pre-pubertal children make an important contribution to our understanding of the relationship between psychopathology and inflammation and the likely direction of effects for two reasons. First, identifying the developmental period at which the association between behavior problems and inflammation becomes apparent has implications for chronic disease prevention strategies. If child psychopathology is truly a risk factor for adult chronic disease via impact on inflammation, interventions aimed at treating or preventing psychopathology in children may reduce risk of chronic disease later in life. Future research is needed to test this hypothesis. For example, child psychopathology interventions could measure inflammatory markers before and after treatment to determine if effective reduction of behavior problems is associated with reducing inflammation. Second, pre-adolescent children tend to have fewer physical health conditions than adults and have not yet initiated many of risk behaviors (i.e., smoking, alcohol consumption) that may influence inflammatory markers. Thus, estimated associations between behavior problems and inflammation in child samples can be expected to have less error, because younger samples have less heterogeneity with respect to other factors that introduce variation into inflammatory markers. Notably, the models had modest R² values in that we accounted for 3–5% of the variance in inflammation, thus suggesting that other factors also contribute to individual variation in CRP and IL-6 levels in children. However, it is important to keep in mind that relatively modest effects can have a meaningful impact when applied to the population-level (Rosenthal and Rubin, 1982). We are unable to compare the size of the R² values to prior studies of behavior problems in relation to inflammation among children prior to puberty (Bartlett et al., 1995; Brambilla et al., 2004; Caserta et al., 2011), because these studies did not present R² values. However, the magnitude of our findings are comparable in size to those obtained in research (using similar models) considering potential effects of childhood adversity on inflammatory markers in youth (Murasko, 2008; Fuligni et al., 2009a).

Our results point to several directions for future research. First, research is needed to identify mechanisms that underlie the observed associations. We did not find evidence for elevated BMI as a pathway and potential behavioral pathways such as smoking or high alcohol consumption are less relevant to pre-adolescent samples. Studies show that behavioral/emotional problems are associated with atypical HPA axis activity in youth (Lopez-Duran et al., 2009; Matthey et al., 2013), and that HPA dysregulation has a broad influence on the immune system (Sternberg, 2001). Therefore, physiological correlates of HPA dysregulation (e.g., abnormal diurnal cortisol profile) may function as potential mediators. Other potential mediators include sleep duration and quality, given that sleep problems are common in children with behavior problems (Alfano et al., 2007; Sung et al., 2008) and poorer sleep predicts inflammation (Simpson and Dingel, 2007). Second, research is needed to examine whether changes in behavior problems are related to changes in inflammatory markers. By incorporating pre-and post-treatment assessments of CRP and IL-6 into trials of behavioral and pharmaceutical interventions for children with behavior problems, investigators could test whether reductions in behavior problems produce reductions in inflammatory marker levels. Such data would provide further support for the hypothesis that behavior problems increase risk for elevated inflammation in youth. Third, studies are needed to clarify whether the timing of onset or duration of behavior problems have implications for later inflammation and long-term disease risk. For example, some research suggests that chronicity of depression (Copeland et al., 2012) or conduct disorder (Odgers et al., 2007) in childhood or adolescence has a marked influence on later CRP. Finally, research is needed to characterize the long-term health risks associated with elevated inflammation in childhood.

Our findings should be considered in the context of several limitations. This study is limited by attrition and incomplete data, and consistent participation was socially patterned such that individuals from more disadvantaged households were less likely to participate. Children with behavior problems were more likely to drop out of the sample (Wolke et al., 2009), therefore this sample underestimates the prevalence of behavior problems in the target population. Other researchers have carried out simulation analyses using the ALSPAC cohort to examine the impact of selective drop-out on the strength of prediction, and have indicated that the validity of regression models to predict behavior disorder is only slightly affected (Wolke et al., 2009). Our models controlled for variables associated with attrition (e.g., household income, maternal education) which may attenuate potential bias (Kleinbaum et al., 1982). However, in light of non-differential missingness, our results cannot be
generalized to the original population-representative cohort, and may underestimate the associations. In addition, our study only had one assessment of inflammation prior to adolescence; therefore, we were unable to examine bi-directional associations. Furthermore, although this study relies on temporally-ordered data, we cannot conclude that behavior problems preceded increased inflammatory markers, as inflammation may have already been up-regulated at the time that the behavior problems were assessed. This limitation also has implications for the tests of mediation; if inflammation was elevated at the time that behavior problems were assessed, paths in the mediation models may be over- or underestimated relative to their true values. Finally, information on asthma or other inflammatory health conditions was not collected at the clinic visit when blood samples were provided. However, we note that results were unchanged when we adjust for past-year reports of asthma or eczema at age 7 (data not shown), suggesting that the observed associations are not confounded by these inflammation-related conditions.

In conclusion, this study provides evidence for associations between behavior problems and elevated CRP and IL-6 in a large sample of children suggesting that associations are already evident prior to adolescence. Specifically, externalizing and somewhat less strongly, internalizing behavior at age 8 were associated with higher levels of CRP assessed two years later, and internalizing and externalizing behavior at age 8 were associated with elevated levels of IL-6 two years later. These associations did not differ for boys and girls, and somewhat surprisingly were not explained by elevated BMI. The findings suggest an etiological pathway that may connect childhood behavior problems to elevated chronic disease risk beginning early in life. Additional research on the connection between childhood behavior problems and inflammation is needed to determine whether behavior problems are causal for elevated inflammation and whether interventions aimed at behavior problems reduce risk of elevated inflammation and, potentially, long term health risks in children.

Role of funding sources
None.

Conflict of interest statement
None declared.

Acknowledgements
We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council (Grant ref.: 74882) the Wellcome Trust (Grant ref.: 076467) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and will serve as guarantors for the contents of this paper. This research was specifically funded by a postdoctoral fellowship to the first author from the Robert Wood Johnson Foundation to support the Early Childhood Innovation Project. The authors would also like to thank Dr. Kate Northstone for her data assistance.

Appendix A. Supplementary data
Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.psyneuen.2013.07.012.

References


C Reactive Protein Coronary Heart Disease Genetics Collaboration, 2011. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ 342, d548.


The Gender Difference in Depression: Are Elderly Women at Greater Risk for Depression Than Elderly Men?

Joan S. Girgus, Kaite Yang, Christine V. Ferri

Abstract: Numerous epidemiological reports have found that adolescent, young adult, and middle-aged girls and women are more likely to be diagnosed with unipolar depression and report greater symptoms of depression when compared to boys and men of similar ages. What is less well-known is whether this gender difference persists into late life. This literature review examines whether the well-known gender difference in unipolar depression continues into old age, and, if it does, whether the variables that are known to contribute to the gender difference in unipolar depression from adolescence through adulthood continue to contribute to the gender difference in the elderly, and/or whether there are new variables that arise in old age and contribute to the gender difference in the elderly. In this review of 85 empirical studies from every continent except for Antarctica, we find substantial support for the gender difference in depression in individuals who are 60 and older. More research is necessary to determine which factors are the strongest predictors of the gender difference in depression in late life, and particularly whether the factors that seem to be responsible for the gender difference in depression in earlier life stages continue to predict the gender difference in the elderly, and/or whether new factors come into play in late life. Longitudinal research, meta-analyses, and model-based investigations of predictors of the gender difference in depression are needed to provide insights into how and why the gender difference in depression persists in older age.

Keywords: depression; gender differences; sex differences; aging; older adults

1. Introduction

It is well-known that females are more likely to be depressed than males, both in terms of unipolar depressive episodes and in terms of depressive symptoms [1,2]. However, while there is a lot of data on this for some age groups, there is much less data for others. For example, it is well-established that there is no gender difference in depression before early adolescence [3], the gender difference in depression emerges between ages of 11 and 15 [4], and that the size of the gender difference increases into adulthood when it apparently levels off [5]. For many years, the accepted wisdom was that the gender difference in depression became smaller and possibly even disappeared altogether in older adults, although there was a general acknowledgement that there was very little data [6]. Since there has been considerable research on depression in older adults in recent years, this seemed to be a good opportunity to review the literature to see the picture of the gender difference in depression in older adults that emerges.
So far, a thorough discussion of the gender difference in unipolar depression in older adults is largely missing from major reviews of the gender difference in depression. There are several reasons for this. First, reviews of the gender difference in depression tend to focus on the emergence of the gender difference during adolescence and subsequently focus more on empirical studies of adolescent, college, and middle adult groups (e.g., [1, 2, 7]). When older adults are included in mixed samples, analysis of gender differences may not be stratified according to age or the sample size of older adults may be too small (e.g., [5]). This results in a limited ability to analyze or interpret data on the gender difference in unipolar depression in older adults in mixed samples. Second, some have expressed uncertainty over whether the gender difference in depression exists in adults aged 55 and older (e.g., [4, 8]). Some uncertainty may come from theorizing and data showing that overall rates of depression and negative affect are attenuated in older cohorts (e.g., [9, 10]), although a decline in the prevalence of depression in older age does not necessarily mean that the gender gap will narrow. Because of the relative absence of studies with data on individuals between the ages of 60 and the end of the life span, the earlier reviews necessarily focused on studies with data from middle-aged and young-old adults, rather than on an elder population.

In this paper, we will review what is known about the gender difference in unipolar depression in old age (age 60 and older). The primary question we will try to answer is whether there is sufficient empirical data to support the existence of a gender difference in depression in older adults. If the gender difference in depression does exist in older adults, we will examine whether the major psychosocial variables that are thought to be responsible, at least in part, for the gender difference in depression in teenagers, young adults, and middle-aged adults play a similar role in depression in older adults, and/or whether there are new psychosocial variables that are responsible for a gender difference in depression in older adults. Finally, we will try to specify the most important research questions that remain to be answered in order to develop a reasonably complete picture of the gender difference in depression throughout the life course.

2. Epidemiology of the Gender Difference in Depression in Older Adults

The first goal was to answer the following question: Is there a gender difference in depression in older adults who are at least 60 years old? We conducted a literature search in March 2017, using the Academic Search Complete, PsycINFO, and MedLine (ProQuest) databases. For each database, the search terms “gender difference”, and “depression”, and “elderly or aged or older or elder or geriatric” were used. This returned 713 abstracts from Academic Search Complete, 3788 abstracts from PsycINFO, and 3155 abstracts from MedLine. Full articles were retrieved and included in our review if they met the following inclusion criteria: empirical research articles published in peer-reviewed journals; participants aged 60 or aged 50 and above with the mean age of the participants at least age 70; both women and men as participants; and, a report of unipolar depression by gender. Lifespan studies were only included in our review if depression data was reported by gender for the group of participants who were 60 and above. Duplicate abstracts were consolidated. These criteria were met by 85 articles that were published between 1980 and 2017. In these 85 articles, the most commonly used measures of depression were the well-validated Center for Epidemiologic Studies Depression scale (CES-D) [11], which was used in 36 (42%) of the studies, and the Geriatric Depression Scale (GDS) [12] which was used in 19 (22%) of the studies. One study used both CES-D and GDS scales [13]. The CES-D and GDS are valid instruments for the assessment of depressive symptoms in geriatric samples [12, 14, 15] The other 29 studies used 24 different measures, including structured clinical interviews, physician diagnoses, and other self-report measures of depressive symptoms (e.g., [16–19]). These 85 studies included participants from 34 different countries, and from every continent except for Antarctica. The vast majority of the articles (72%) were published in the last 10 years, reflecting the increased interest in the subject of depression in the elderly in recent years.

In 69 (81%) of the 85 studies, older women, as compared to older men, had a significantly greater likelihood of a depression diagnosis based on clinical interviews or cut-off scores, or significantly
more depressive symptoms on standard measures. Of the 69 studies, four found a significant gender difference in some groups but not others. In 14 of the 85 studies (16%) there was no significant difference in depression scores or diagnoses between women and men. In two of the 85 studies (2%), men had significantly greater depression scores than women. Taken altogether, this answers the central question of this paper: the gender difference in depression, with women more likely to be depressed than men, is found in people over the age of 60. (See Table S1 in Supplemental Materials for the full list of articles reviewed, e.g., [20–70]).

Only 16 (18%) of the 85 studies that met the criteria for inclusion in this review did not find a significant gender difference in depression, in which women scored higher than men. In these 16 studies, the characteristics of the samples and the ways in which depression was assessed did not seem to differ in any significant way from the 69 studies in which a gender difference in depression was found. Furthermore, these 16 studies reported data on samples from 11 different countries (France, Italy, Sweden, Taiwan, United States, Australia, Germany, Brazil, Singapore, China, South Korea).

In summary, 69 of 85 (81%) studies that reported on the gender difference in depression in the elderly found that older women were more likely to have a diagnosis of depression or a greater number of depressive symptoms when compared to older men. Although interpretation of the epidemiological findings is made somewhat more difficult because only a comparatively small percentage (29%) of the studies excluded participants based on conditions such as dementia, cognitive function, chronic illness, disability, and co-morbid psychological disorders, the consistency of the basic finding of a gender difference in depression in the elderly extends the consistent epidemiological finding that there is a gender difference in depression among adolescents, younger adults, and middle-aged adults (e.g., [1,71]).

3. Gender Differences in Depression across the Lifespan

Some empirical studies have asked whether the gender difference in depression changes across development stages from adolescence and young adulthood into late life. Our screening process returned 10 studies with participants from 18 different countries that addressed the question of whether the gender difference in depression changes in magnitude across the life span. The results of these analyses were mixed. Three of the studies reported that the gender difference in depression, while remaining significant, was narrower in elderly adults as compared with young- and middle-aged adults [72–74]. Three studies reported that the size of the gender gap in depression increased with age [75–77]. Finally, four studies found that the gender difference in depression remained the same across age groups [78–81].

One possible reason for the mixed results is the differences in the age groups that were analyzed in the various studies. In the studies that found no change in the gender difference in depression over age, one compared all of the participants under age 55 with all of the participants over age 55 [79], a second compared all of the participants under age 50 with five, five-year cohorts over the age 50 [80], and a third compared of the participants over age 55, divided into eight, five-year, cohorts [78]. In other words, in these three papers, most of the data were drawn from participants over the age of 50. The fourth study that found no difference in the gender difference with age compared a small group of 18-29 year olds to a small group of 50–76 year olds [81]. In the studies that found a decrease in the gender difference in depression with age, the participants were spread more completely throughout the life span. One compared a sample of 20–35 year-olds to a sample of 70–85 year-olds [73]; a second compared samples of three age groups, ages 25–35, 45–55, and 65–75 [72]; and, the third compared a slightly different set of three age groups, ages 20–24, 40–44, and 60–64 [74]. Of the three studies that found an increase in the gender difference in depression with age, one had participants who ranged in age from 45 to 85, divided into four, 10-year cohorts [77], one had participants who ranged in age from younger than 30 to 80 and older, divided into seven cohorts [76], and the third had participants divided into 15-year cohorts, beginning with age 18 and concluding with a cohort of over 65 year-olds [75].
In general, the studies that explored the possibility of changes in the gender difference in depression across the life span used a cross-sectional approach, with different participants at different ages. As a result, cohort effects cannot be ruled out. Large-scale epidemiological studies have revealed that a cohort effect is present for depression, with more recent cohorts reporting higher rates of depression when compared with older cohorts [9,10,82]. It is not clear, however, that higher rates of depression in more recent cohorts would necessarily affect the size of the gender difference in depression. Three of the studies combined longitudinal data gathering with a cross-sectional design [74,76,80]. Nguyen & Zonderman (2006) and Mirowsky (1996) followed their participants over a 6–10 year period, and Leach et al. (2008) followed their participants over a 20-year period. One of these studies found an increase in the gender difference in depression with age [76]; one of these studies found a decrease in the gender difference in depression with age [74]; and, the third study found no change in the size of the gender difference in depression with age [80]. The particular age cohorts analyzed differed substantially among these three studies and so the lengths of time and the ages included in them may not have been sufficient to test the question of whether there are changes in the gender difference in depression as people enter and move through old age.

All in all, the currently available data are clearly mixed on whether and how the gender difference in depression changes across the life span. There is no readily perceivable pattern based on the sampling procedures (or the depression measure) used in those studies that have tried to examine this question.

Two studies examined the incidence of the initial depressive episode between the ages of 18 and 65, and asked whether more women than men experienced an initial depressive episode at each age [83,84]. In both of the studies, the probability of an initial episode of depression was higher for women than for men across this age range. However, the probability of an initial episode of depression declined with age for both women and men, with the gap between them narrowing as the probability became extremely low for both genders in old age. On the other hand, in a study of initial depressive episodes in a longitudinal sample that was tested every five years between the ages of 70 and 85, Palsson, Ostling, and Skoog (2001) found that initial depressive episodes increased across this age range for both men and women, and women were consistently more likely than men to experience an initial episode [85].

4. Psychosocial Predictors of the Gender Difference in Depression

While the psychosocial predictors of depression have been extensively studied in people of all ages, the psychosocial predictors of the gender difference in depression have been studied much less, and almost not at all in older adults. What research there is on the psychosocial predictors of the gender difference in depression has generally focused on adolescents and young adults, and utilized diathesis-stress models. In this research, the diatheses are vulnerabilities that make individuals more susceptible to depression when stressors occur in their lives, and the question is which stressors and which vulnerabilities are more prevalent in the lives of females than in the lives of males [86]. The research on the psychosocial predictors of the gender difference in depression in older adults, however, has generally not utilized any underlying model, but instead has usually looked at single variables that might be related to the gender difference in depression. While sometimes multiple variables are included in a single study, the question being asked is almost always which of these variables accounts for the most variance in the gender difference in older adults.

In addition, many researchers have looked separately at which variables predict depression in older women and which variables predict depression in older men. These studies rarely ask whether there are gender differences in the psychosocial variables that predict depression in each gender (and whether any gender differences in those variables might lead to women being more negatively affected than men). There are very few studies that look at whether there is a gender difference in a predictor variable that predicts the gender difference in depression. This is unfortunate because only with
such studies can we conclude that that variable can, at least in part, explain the gender difference in depression in the elderly.

In this section, we will examine the results of the studies that have looked at gender differences in psychosocial variables that might predict the gender difference in depression in older adults. First we will consider whether there is evidence that the variables that predict the gender difference in depression in adolescents and younger adults might also predict the gender difference in depression in older adults, and then we will consider whether there are additional variables that arise in old age that might be responsible at least in part for the gender difference in the elderly.

This section combines studies that were done in many different countries. In some studies, data from several different countries were examined. The results were quite consistent across countries, and so the results are reported without considering the countries in which the data were gathered.

5. Psychosocial Predictors of the Gender Difference in Depression That Occur throughout the Life Span

In adolescents and younger adults, the predictors of the gender difference in depression that have been identified include: stressors, coping styles (e.g., rumination); depressogenic personality styles (e.g., sociotropy); social support; the physiological changes of puberty; gender role intensification, sexual abuse and assault; and, negative body image [2,7,86]. The physiological changes of puberty and gender role intensification have been viewed as specific to adolescence and young adulthood, while negative body image and sexual abuse and assault have been seen as more prevalent in adolescence and young adulthood than in later life stages. Coping styles, depressogenic personality styles, and some stressors, on the other hand, would seem to be candidates for predictors of the gender difference in depression throughout the life span. In this section, we will explore whether these variables seem to predict the gender difference in depression in the elderly, as they do the gender difference in depression in adolescents and young adults.

5.1. Stressors/Negative Life Events

There are many different kinds of stressors. The most commonly measured stressors in the depression literature generally are negative life events that occurred within the previous year. However, very few studies have asked whether there are gender differences in negative life events in the elderly that are related to the gender difference in depression in this age group, despite the fact that negative life events occur frequently in the lives of the elderly. In the few studies that have examined this question, it is not clear whether older women experience more negative life events than older men: in two studies they did [87,88] and in two studies they did not [73,89]. Chan, Kwok et al. (2012) found that elderly women had significantly more depressive symptoms than elderly men, but did not analyze their data to see whether the gender difference in negative life events predicted the gender difference in depression [88]. Seematter-Bagnoud et al. (2010) did not analyze their data either for a gender difference in depression or for a possible relationship between the gender difference in negative life events and a possible gender difference in depression [87].

5.2. Coping Styles

Cognitive coping skills are commonly considered to be predictors of the gender difference in depression in adolescents and younger adults ([90], for meta-analysis, see [91]). The most extensively researched coping style that has been linked to the gender difference in depression is rumination, which was initially proposed by Susan Nolen-Hoeksema. Rumination occurs when stressors lead to negative moods and individuals respond to the negative moods by dwelling on them, their causes, and their implications, rather than engaging in either problem-solving or distraction [92]. Women ruminate more than men when faced with a stressor [93], and this gender difference in rumination is well established as a risk factor for the gender difference in depression in adolescents and adults [94]. Only a few studies have examined rumination in old age [72–74,81]. These studies all compared older
adults with younger ones. Collectively, they found that rumination declines with age, but at all ages, women ruminate more than men. Those studies that examined whether rumination mediates the relationship between gender and depression found that it did [72,73,81].

There is a substantial history of dividing coping styles into approach styles (e.g., problem-solving) and avoidance styles (e.g., avoiding thinking about a problem) [95]. Two studies have examined gender differences in coping styles in the elderly [96,97]. Both found that elderly women use avoidance coping styles more than men do, and Lutzky and Knight (1994) found that this gender difference was related to the gender difference in depression [96]. In these two studies, women also used approach coping styles (especially seeking social support) more than men did, but only Lutzky and Knight (1994) tested whether this gender difference was related to the gender difference in depression, and they found that it was not [96].

5.3. Interpersonal Orientation (Sociotropy/Dependency)

Some depressogenic personality characteristics that are quite commonly assessed in younger samples have not been examined in older samples. For example, dependency and sociotropy are interpersonal orientations that have been proposed to contribute variance to the gender difference in depression in younger samples [98–100]. This review of the gender difference in depression did not find any articles that examined dependency and sociotropy as predictors of the gender difference in depression in the elderly. While we have some preliminary data that suggests that sociotropy declines with age [101], it is unclear whether the gender difference in sociotropy that is still present in old age explains some of the gender difference in depression in the elderly.

5.4. Social Support

On the face of it, social support looks like a potentially important variable when considering the predictors of depression. In general, social support is seen as buffering against depression; that is, it is assumed that social support protects a person against depression [102]. Insofar as social support is viewed as a buffering variable, the obvious hypothesis would be that it is the absence of social support that predicts depression. Therefore, if a gender difference in social support plays a role in the gender difference in depression in the elderly, elderly women would be expected to have less social support than elderly men. Rather, there is significant evidence that older women have higher rates of social contact, support, and participation, and more extensive social networks, than older men [89,103–107]. Some research shows similar quality and quantity of social relations in elderly men and women [108,109], and other studies have found that elderly men had higher perceived level and perceived availability of social support than elderly women [110,111]. Even within a single study, there are sometimes results that could be seen as contradictory. For example, van Grootheest et al. (1999) found that elderly women reported receiving significantly more emotional support and significantly less instrumental support than elderly men [105]. Leach et al. (2008) reported more positive support and fewer negative events from friends for elderly women as compared to elderly men, but also reported no difference between elderly women and men in positive support and negative events from family [74]. Lee & Lee (2011) reported that older women visited more frequently with friends than older men did while older men visited more frequently with their children than older women did [112].

One of the reasons for the mixed results may be that there are many different ways of defining social support, and it is difficult to figure out whether there is any comparability among them. For example, social support can be either instrumental or emotional. It can be defined as the size of a person’s social network or as whether a person has a best friend or confidante. It can be seen as coming from family or from friends. One or more of these kinds of social support could have more of an effect on one gender or the other, and these effects could be positive or negative. For example, it is possible that elderly women are more vulnerable to the absence of social support than elderly men and that this difference in vulnerability could account in part for the gender difference in depression.
6. Psychosocial Predictors of the Gender Difference That Are More Prevalent in Older Adults

The psychosocial variables that are more prevalent in old age than in younger age groups can best be characterized as new stressors that are more likely to arise or intensify in old age. These include: widowhood/living alone; illness; cognitive decline; financial strain/poverty; and, caregiving. Each of these variables increases in old age, and the question then becomes: are there gender differences in these variables in old age and, if so, do these gender differences predict the gender difference in depression in old age?

6.1. Widowhood/Living Alone

Marital status and living arrangements are the most studied psychosocial variables in the literature on depression in the elderly. These are not exactly the same variables although they certainly overlap; for example, it is possible for an elderly person who is widowed or divorced to be living alone or to be living as part of an extended family. We will first examine the literature on marital status, and then we will explore the literature on living arrangements.

The data consistently show that elderly women are less likely to be married and more likely to be widowed than elderly men [13,74,78,87,90,105,107,113,114]. Although the literature makes it clear that elderly adults who are married have less depression than elderly adults who are widowed/divorced/separated [107,115–118], there is much less data on whether the gender difference in marital status is related to the gender difference in depression. Since women are more likely than men to be widowed/divorced/separated and since people who are widowed/divorced/separated have more depressive symptoms than people who are married, it would not be surprising if marital status were a predictor of the gender difference in depression, but direct tests of this have been infrequent, and have not supported this view. Rather, what little data there are suggest that women who are married are more depressed than men who are married, but there is not a gender difference in depression in older adults who are separated/divorced/widowed [74,105,114,117]. Some investigators have focused specifically on widowhood as a life event that predicts geriatric depression. In general, these studies found no gender differences in depressive symptoms after widowhood [111,118].

Other investigators have proposed that a gender difference in living alone might be a predictor of the gender difference in depression. Many more elderly women live alone than elderly men [13,87,116]. However, elderly men who live alone have more depressive symptoms than elderly women who live alone [116,119], exactly the opposite of what one would expect if a gender difference in living alone were a predictor of the gender difference in depression.

6.2. Poor Health/Chronic Illness

Health status seems to be an obvious variable to examine in relation to the gender difference in depression in the elderly since health tends to decline as people age. Health can be measured in a number of ways: number of chronic conditions, ability to do the tasks of daily life, and self-ratings of health are the approaches that have been used most frequently in studies of the elderly.

The first question, however, is whether there is a gender difference in health. Do elderly women have poorer health and more chronic conditions than elderly men? In general, the answer to this question seems to be yes. There is a gender difference in health status in the elderly, with women experiencing more chronic conditions, demonstrating less ability to perform daily tasks, and subjectively reporting poor health [13,74,78,87,106,112,120]. Given this, health status becomes a candidate for predicting the gender difference in depression in the elderly. Only one study has asked this question directly. Noh et al. (2016) found that there was no difference in depression scores between men and women who were not diagnosed as having a disability but women who were diagnosed as having at least one disability were significantly more depressed than men who were diagnosed as having at least one disability [120].
6.3. Cognitive Decline/Dementia

The studies of the relationship between cognitive decline/dementia and depression in the elderly differ from studies of the relationship between other variables and depression in the elderly in an important way. Whereas, studies of other variables generally ask whether the variable predicts depression, studies of cognitive decline/dementia ask whether depression predicts cognitive decline. While the reason(s) for this is not explicitly discussed in the literature, it seems likely that it reflects the difficulty of assessing depression in individuals with dementia. There are considerable data that older women are not only more likely to be depressed than older men, but also are more likely to be diagnosed with dementia, especially over age 85 [121–124], although some studies suggest that this gender difference is not found in North America and Australia [124,125]. Yet, despite the fact that elderly women seem more likely to be diagnosed both with dementia and with depression than elderly men, the data do not support the idea either that the gender difference in depression is related to the gender difference in dementia.

The data on whether depression is a risk factor for or a correlate of cognitive decline are mixed, and gender differences have rarely been reported. In the studies reviewed by Sevick, Rolih, and Pahor (2000), the data on depression as a predictor or correlate of dementia or cognitive decline were mixed, with no gender differences being reported [126]. In the meta-analyses conducted by Jorm (2000), there appeared to be a small but significant association between a history of depression and subsequent dementia and cognitive decline over periods of less than a decade, but again no gender differences were reported [127]. In a study of individuals with dementia, elderly women with mild dementia had more depressive symptoms (on a scale completed by a caregiver) than elderly men with mild dementia, but there was no gender difference in caregiver rated depressive symptoms in individuals with moderate dementia [128]. Overall, then, it is not clear whether a history of depression is a risk factor for dementia and cognitive decline, and the data generally suggest that there are no gender differences in whatever relationship there is between depression and cognitive decline.

6.4. Financial Strain/Poverty

Although financial strain and poverty are thought to be more prevalent in women than in men throughout the life span, perhaps surprisingly, we found very few studies in the depression literature that reported whether elderly women experience more financial strain and poverty than elderly men. Those studies that have examined whether elderly women are more likely to be poor than elderly men have found this to be the case [129,130]. Most of the studies that have asked whether financial strain or poverty predicts depression in the elderly have asked this question separately for men and women [130–135]. There do not seem to be any studies that have asked whether a gender difference in financial strain or poverty predicts the gender difference in depression in old age.

6.5. Caregiving

Caregiving is by no means unique to older adulthood, but it is certainly a demand placed on many older adults. Pinquart and Sorenson (2006) conducted a meta-analysis that found higher rates of depressive symptoms in female caregivers than in male caregivers, with women reporting higher levels of burden, more hours of care, more caregiving tasks, and more personal care than male caregivers [108]. This meta-analysis combined spouses, children, and other caregivers so, although those needing care were all elderly, the caregivers were not. There were several other findings of interest: the gender difference in depression was larger for older caregivers as compared to younger, and for samples of caregivers as compared to samples that were not selected for caregiving. Not surprisingly, this meta-analysis found gender differences in several stressors that were specific to caregivers (e.g., hours of care per week), although, somewhat unexpectedly, the meta-analysis did not find any gender difference in either formal or informal support. After controlling for stressors and social support, the meta-analysis found that the gender difference in depression in caregivers was no longer
larger than the gender difference in depression in a sample not selected for caregiving, supporting a stress and coping model of the additional gender difference in depression found in caregivers.

The Pinquart and Sorenson (2006) meta-analysis examined studies in which those receiving care might be debilitated in various ways [108]. Other studies have focused on the caregivers of those with dementia or Alzheimer’s disease [110,136]. Although these studies did not directly test whether there were caregiver variables that predicted the gender difference in depression in the caregivers, they did find that female caregivers had higher subjective burden and perceived stress than male caregivers.

7. Conclusions

The present review found strong empirical evidence for the gender difference in depression among older adults aged 60 and above. Older women scored higher than older men on dimensional measures of depressive symptoms and older women have higher rates of diagnosis of unipolar depression compared to older men. This pattern was observed in 69 (81%) of 85 studies. The majority of studies included in this review assessed depression using dimensional self-report instruments of depressive symptoms. Although self-report instruments, such as the CES-D and GDS, are not the same as a diagnostic clinical interview, they nevertheless are valuable tools for depression screening and assessment of symptom severity, are highly predictive of well-being, and are valid assessments in older adults [12,14,15].

Although subclinical levels of depressive symptoms are more commonly experienced in geriatric samples than major depressive disorder [137], subclinical depressive symptoms contribute substantially to a decreased quality of life, including cognitive decline, longer hospital stays, and poorer health in late life [137]. Therefore, the substantial evidence for the existence of a gender difference in depression, even at subthreshold levels, is a consequential finding.

The literature reviewed for this paper was impressive in its geographic and cultural diversity, with every continent except for Antarctica being represented in the papers reviewed. The majority of the papers reported data from North America, Asia, and Europe, with fewer studies taking place in South America, Africa, and Australia. Population-based samples predominated in the studies reviewed, with remarkably little dependence on convenience samples. Taken altogether, the existence of a gender difference in depression in the elderly, with women being more likely to be depressed, appears to be a very robust finding.

While the existence of the gender difference in depression in the elderly seems to be solidly established, we know much less about why this gender difference occurs in the elderly, the magnitude of this effect over ages, or whether this effect occurs in the elderly for the same reasons that it occurs at earlier life stages. In particular, we were unable to ascertain whether the data were similar for older adults (age 60–80) and for very old adults (older than age 80). Those studies that explored the gender difference in depression across a broader age range in order to examine whether the gender difference in depression changed in magnitude across the life span provided mixed data, perhaps because most of these studies used different age ranges [72–81]. However, even among those studies that included cohorts of participants ranging in age from the late teens to over 65, the results were mixed: three found an increase in the gender difference in depression in later cohorts [72–74], and three found a decrease in the gender difference in depression in later cohorts [75–77].

In general, the vast majority of the studies that explored predictors of depression in the elderly looked at what predicted depression in women and men separately. There were relatively few studies that reported gender differences in potential predictor variables, and an even smaller number of studies that actually tested whether the gender differences in predictor variables accounted for a significant part of the variance in the gender difference in depression. In addition, most of the studies focused on stressors that increase dramatically in old age, such as widowhood/living alone, poor health/chronic illness, cognitive decline, financial strain/poverty, and caregiving. There were relatively few studies with elderly participants that explored predictors, such as negative life events, coping styles, and interpersonal orientation, which have been extensively researched in adolescents and adults.
The only predictor of the gender difference in depression that has been studied extensively throughout the life span is social support, but the data on the role of social support are very mixed and difficult to interpret, probably because there are so many different ways to define social support (and its absence).

There does seem to be some indication that dysfunctional coping styles, such as rumination and avoidance coping, are more prevalent in elderly women than in elderly men [72–74,81,96,97], just as they are in adolescents, young adults, and middle-aged adults ([72,138], for meta-analysis, see [94]). However, considerably more research will be needed in order to understand the role that these coping styles may play in the gender difference in depression in the elderly. Research is also needed on the question of whether interpersonal orientation (sociotropy/dependency) plays a role in the gender difference in depression in the elderly as it apparently does at younger ages [98–100].

Although particular stressors—widowhood/living alone, poor health/chronic illness, cognitive decline, financial strain/poverty, and caregiving—have been extensively studied in the elderly, the broader range of negative life events that have been studied in adolescents and adults, and that generally represent the stressors in diathesis-stress models of depression, are notably absent from the literature on gender differences in old age. More research on the broad range of negative events in the lives of the elderly might prove informative for understanding the gender difference in depression in this age group.

Although we know more about the particular negative events of old age than we do about negative life events in general, we do not know nearly enough about any of them to be able to specify whether a particular negative event makes a contribution to the gender difference in depression in the elderly. People who are not married (widowed, divorced, or separated) are more likely to be depressed than people who are married, and elderly women are much less likely to be married than elderly men [107,115–118]. However, direct tests of not being married as a contributor to the gender difference in depression are scarce, and the few studies that have looked at this have found that elderly women who are married are more depressed than elderly men who are married, but there is no difference in depression between women and men who are widowed, divorced, or separated [74,105,114,117]. In a major meta-analysis, Pinquart and Sorenson (2006) found clear evidence that female caregivers reported more depression than male caregivers [108]. However, the studies in this analysis included caregivers of a wide range of ages, while those receiving care were all elderly. Nor did this meta-analysis test whether being a caregiver of an elderly person accounted in part for the gender difference in depression in the caregivers. Clearly research that focuses on the elderly caregivers of elderly care recipients is needed before there will be a clear picture of the role that caregiving might play in the gender difference in depression in the elderly. Poor health and chronic illness, cognitive decline, and financial strain and poverty need even more research before they can be considered possible contributors to the gender difference in depression. What data there are suggest that elderly women suffer from poorer health in general than elderly men [13,74,78,87,106,112,120], but we know very little about whether poorer health is related to the gender difference in depression. Elderly women are apparently more likely to be poor than elderly men [129,130], but again the connection between this and the gender difference in depression is obscure. Gender differences in cognitive decline and dementia, and their possible relationship to the gender difference in depression, are even more difficult to examine, largely because it is so difficult to assess depression in individuals with dementia. As a result, most of the research on the question of the relationship between cognitive decline and depression has asked whether depression is a risk factor for cognitive decline rather than whether cognitive decline is a risk factor for depression [126–128].

It seems clear that there is a new set of stressors that arise in old age. Unfortunately, there is not sufficient research to be certain whether any of these new stressors contributes to the gender difference in depression in the elderly. There is, however, a hint in the literature that the risk factors for a depression diagnosis in the elderly may be different from the risk factors at younger ages. Studies have documented a U-shaped function for the gender difference in initial diagnosis of depression between the ages of 18 and 65, with essentially no gender difference at the youngest and oldest ages [83,84].
However, a study that examined the gender difference in the initial diagnosis of depression in a sample of individuals aged 70–85, found that initial depressive episodes increased for both women and men across this age range with women being consistently more likely than men to experience an initial episode [85]. This strongly suggests that a new set of risk factors, which are more prevalent in women than in men, enters the lives of individuals after age 65.

Severe and moderate (but not mild) depression is also associated with increased mortality in older adults [139,140], and chronic depression is associated with an increased mortality in the elderly [139]. The data on gender differences in the relationship between depression and mortality are mixed, however. Although elderly women are more likely to be depressed than elderly men, some studies found that men who are depressed have a higher risk of mortality than women who are depressed [139,141]. At least one other study, however, found an increased risk of mortality for depressed women but not for depressed men [142]. Older men are also at higher risk for death from suicide than older women, and the reasons for suicide in older adults also differ by gender [143].

7.1. Recommendations for Future Research and Clinical Application

We designed this review of the literature on the gender difference in depression in the elderly to examine, first, whether the well-known gender difference in depression among adolescents, young adults, and middle-aged adults, was also consistently found in the elderly and, second, to examine what was known about risk factors for depression in the elderly. The present review establishes the existence of the gender difference in depression among the elderly and reveals important gaps in the literature about what factors are responsible for this gender difference. Below, we describe ways to address these gaps through empirical research. We also briefly touch on applications of these findings for clinical practice, especially as the applications pertain to depression screening of senior populations.

7.2. Theory-Based Approaches to Understanding the Gender Difference in Depression

Although the general literature on the gender difference in depression, which is almost entirely based on samples of adolescents and young adults, has examined and compared various models designed to explain the gender difference [83], model-driven approaches to understanding the gender difference in depression in elderly adults are noticeably rare. In the literature on the etiology of the gender difference in depression in adolescence, three models have been proposed to explain why girls are more likely than boys to experience depression: the diathesis-stress model; the mediation model; and the transactional model. The diathesis-stress model states that the interaction between stress and certain personality vulnerabilities increases risk for depression, with females being more likely than males to experience stressors [144], and have depressogenic personality diatheses [145]. Under the mediation model, different ways of coping with stressors trigger depression, and males and females differ in their tendencies to use more or less depressogenic coping styles [146]. The transactional approach emphasizes the cyclical nature of depression and proposes that individuals with depression engage in behaviors that perpetuate symptoms [147,148], with individuals with depression, more of whom are female, more likely to engage in such behaviors.

The emergence of the gender difference in depression in adolescence is likely overdetermined: there is empirical support for many of the variables, as well as all of the models that have been proposed [86]. Given this, it seems reasonable to assume that the gender difference in depression is overdetermined throughout the life span. Thus, in order to understand the lifespan development of the gender difference in depression, we recommend research approaches that combine or compare multiple models (e.g., [90]). We also recommend the examination of variables and hypotheses that predict the emergence of the gender difference in adolescence and explain the gender difference in depression in young adults. Only by doing this, will it be possible to determine whether the gender difference in depression in the elderly is predicted by the same factors that contribute to the gender
difference at younger ages. This may reveal important developmental processes that are involved in the onset and perpetuation of gender differences in depression.

7.3. Meta-Analytic Research

We strongly recommend that a comprehensive meta-analytic review of the empirical literature on the gender difference in depression, including the empirical literature on the gender difference in older adults, be conducted as the next step towards understanding gender and depression over the lifespan. Some meta-analyses have already been conducted on subsets of older adults (e.g., [108]). However, a large-scale meta-analysis that includes younger and older samples would reveal whether the magnitude of the gender difference in depression changes over the life span. In the present review, we found mixed evidence for the shape of the gender difference over the life span, with studies being nearly evenly split between findings that the gender difference narrows in older cohorts, widens in older cohorts, or stays the same across age groups. Only a meta-analytic review would be able to quantify the effect size of the gender difference in depression across multiple studies of the lifetime prevalence of depression. We recommend that the predictors of the gender difference in depression that are discussed in the present review be examined as potential moderators of the gender difference in older adults.

7.4. Gaps in the Literature

A critical issue that needs to be addressed in future research on gender, depression, and aging, is stratification by age group in individuals over the age of 60. Most studies included in the present review lumped all of the individuals over the age of 60 together in their analyses. This can be contrasted with research on adolescents and young adults where narrow cross-sectional cohorts are carefully defined. In addition, given the expected life span, at least in developed countries, individuals at age 60 can expect to live for approximately 20 more years. Thus, research that combines together participants over age 60 would be similar to research that combines individuals between the ages of 20 and 40. There are important changes in health, ability, and marital status that take place in the decades following age 60. Ideally, research on gender differences and depression in older adults should stratify findings by age in order to examine the correlates of developmental changes that are particular to different decades. Future studies on gender differences in late life depression should stratify by age, examine interactions between gender and age, and identify predictors that particularly affect individuals in different age categories.

The studies in the empirical literature that examined the gender difference in depression across large parts of the life span used cross-sectional designs [72,73,75]. Such designs can confound inter-individual with inter-cohort differences. Thus, we recommend that studies adopting this approach should take care to isolate factors that predict intercohort differences. Seedat and colleagues (2009) found that the gender difference in depression has narrowed in more recent cohorts, and that changes in traditional gender roles may account for this difference [75]. However, others have found the opposite effect, where the gender difference in depression is smaller in older cohorts than in more recent cohorts of younger adults [72,76]. Cohort-level changes in mental health awareness and stigma, the acceptance of negative affectivity, and developmental changes in coping and sociomemotional selectivity have all been proposed to account for this pattern [9,72,149]. Ideally, studies that stratify by cohort should examine multiple factors that represent possible intercohort differences, and at the same time, could contribute variance to the gender difference over time.

Some of the studies included in this review followed multiple cohorts longitudinally for a number of years, but none of them followed the cohorts long enough to collect life span developmental data (e.g., [80,111]). It is obviously extremely difficult to follow a single cohort across the life span. So, it is not surprising that studies reporting on the gender difference in depression have generally not tried to use this approach. It would be useful to have studies that combine a cross-sectional with a longitudinal approach. For example, a study that began with cohorts sampled every 10 years across the life span,
and then continued with yearly follow ups over the subsequent 10 years, would make it possible to separate inter-individual from inter-cohort differences.

Using self-report data with elderly participants, several studies have found that the gender difference in first-onset of depression follows an inverted U-shaped function across the lifespan, with the gender difference converging in adolescence and around age 65, and widening from adolescence to middle adulthood [83,84,150]. Zisook and colleagues’ (2004) data show that women are more likely than men to experience early initial episodes of depression, which are linked to more severe and enduring depression [150]. Longitudinal research would be useful in identifying the trajectory of depression after first onset, and whether the gender difference in the early onset of depression accounts for some of the variance in the gender difference in depression at least until age 65.

The majority of studies included in the present review did not report on screening for comorbid psychological disorders, dementia, or health problems, and those that did screen for these issues usually did not exclude participants based on that screening. As a result, it is difficult to determine whether the reported gender difference in depression is due, at least in part, to other health-related factors. Older adults may present with depression as a symptom or as a consequence of an underlying health problem or cognitive decline [151–153]. We recommend that future studies use methods that provide better control of health variables.

7.5. Relevance to Clinical Practice

The higher prevalence of depression in women than in men highlights the need for mental health professionals to target women in depression screenings and interventions. Women’s health initiatives would benefit from including depression as a target for educational, prevention, and treatment interventions. Since elderly men who live alone are also at a high risk for depression, they could be targeted by primary care physicians for depression screening during medical visits or in community-based depression screening programs.

Increasing our understanding of the predictors of the gender differences in depression can help practitioners select more appropriate interventions. With women who are ruminating and using avoidance coping skills, cognitive-behavioral interventions might be best suited to treat depression. Interpersonal psychotherapy would also be an appropriate intervention for individuals for whom social support, or the lack thereof, is impacting depressive symptoms. Since women who are married are more susceptible to depression than men who are married, couple interventions may be more appropriate for married women than married men.

7.6. Final Thoughts

Depression in older adults is a serious concern. The experience of depression can lead to personal distress and disability, and pose a significant impediment to successful aging. Generally speaking, successful aging is a multidimensional construct that includes psychological health, physical health, social support, and the ability to carry out the tasks of everyday life [154]. Unsurprisingly, intercorrelations exist among the determinants of well-being in the elderly. Our review finds that the gender difference in depression is present in elderly samples and is similarly overdetermined. Multiple predictors account for the gender difference in depression in the elderly in complex and sometimes contradictory ways. Elucidating the interactions between the predictors and the lifespan development of the gender difference in depression will likely contribute to the timely prevention, identification, and treatment of depression in the elderly.

Supplementary Materials: The following are available online at www.mdpi.com/2308-3417/2/4/35/s1, Table S1: Articles reporting the gender difference in depression in older adults.

Acknowledgments: We gratefully acknowledge the funding support of the Stockton Center on Successful Aging Faculty Research Scholar Award.

Conflicts of Interest: We declare no conflict of interest.
References


86. Girgus, J.; Yang, K. Gender and depression. *Curr. Opin. Psychol.* 2015, 4, 53–60. [CrossRef]


117. Zhang, B.; Li, J. Gender and marital status differences in depressive symptoms among elderly adults: The roles of family support and friend support. *Aging Ment. Health* 2011, 15, 844–854. [CrossRef] [PubMed]
Anstey, K.J.; Burns, R.A.; Birrell, C.L.; Steel, D.; Kiely, K.M.; Luszcz, M.A. Estimates of probably dementia prevalence from population-based surveys compared with dementia prevalence estimates based on meta-analyses. BMC Neurol. 2010, 10, 62. [CrossRef] [PubMed]


© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
Social stress, obesity, and depression among women: clarifying the role of physical activity

Karen D. Lincoln

Suzanne Dworak-Peck School of Social Work, University of Southern California, Los Angeles, CA, USA

ABSTRACT

Objective: This study examined the role of stress in the association among physical activity, obesity, and depression among women. The extent to which physical activity moderated these relationships was also examined.

Design: Data from the National Survey of American Life (N = 3235) and multivariable regression analyses were used to examine the effects of chronic stress, material hardship, racial discrimination, and physical activity on obesity and depression among African American, Caribbean Black and White women.

Results: Stress was not related to body mass index (BMI) for African American or White women, but chronic stress was associated with higher BMI for Caribbean Black women. Stress was associated with depressive symptoms, but there was variation by the type of stressor under consideration. Physical activity was associated with fewer depressive symptoms and lower BMI, but the relationships varied by type of stressor and race/ethnicity. Physical activity moderated the effect of chronic stress on depressive symptoms and BMI, but only for African American women who reported high levels of chronic stress. Among White women, physical activity moderated the effect of racial discrimination on BMI for those who reported experiencing both high and low levels of discrimination.

Conclusion: This study was the first to document physical activity as a moderator in the relationship among stress, depression, and obesity using a nationally representative sample of racially/ethnically diverse women. Findings provide insight into the role of stress in relation to depression and obesity while highlighting heterogeneity among Black Americans.

ARTICLE HISTORY

Received 4 December 2016
Accepted 14 June 2017

KEYWORDS

Depression; obesity; stress; African Americans

Introduction

Black women have the highest prevalence rates of obesity compared to any other race or gender group in the United States. In 2014, 56.9% of Black women aged 20 years or older were obese, compared with 35.5% of White women (Ogden et al. 2015). If current trends in obesity persist, it is estimated that all Black women will be overweight or obese by 2034 (Wang, Beydoun, and Liang 2008). Many researchers have identified obesity as the main factor driving racial disparities in health among women.
Although interventions focusing on diet and exercise can be effective, they result in minimal reduction in body mass index (BMI) and are reportedly less successful for Black females. In a study of 1148 adolescent girls, higher levels of physical activity were prospectively associated with lower levels of obesity in White adolescent girls but there was no evidence of an association between levels of physical activity and obesity in Black girls (White and Jago 2012). Another study found that Black women demonstrated significantly less weight loss than other racial and gender groups in intensive lifestyle interventions (West et al. 2008). Authors of this study concluded that the lack of success among Black women may have been tempered by powerful and ambient environmental and sociocultural factors that have not yet been identified or understood.

**Stress response, obesity and depression**

Because the social conditions in which people are born, live, work, and age are the most important determinant of their physical and mental health status, it is important to consider factors related to the environmental and social contexts that produce or perpetuate individual behaviors related to obesity, such as social stress. The association between stress, physical and mental health differs markedly by race and ethnicity (Lincoln and Chae 2012; Lincoln and Takeuchi 2010; Lincoln et al. 2010; Woodward et al. 2012), due in part to differences in exposure to social and environmental stressors (Evans-Campbell, Lincoln, and Takeuchi 2007; Keith et al. 2010; Lincoln, Chatters, and Taylor 2003; Lincoln et al. 2007; Turner and Lloyd 2004; R. Turner and Lloyd 1999), the extent to which experiences are appraised as stressful (Chae et al. 2010; Chae, Lincoln, and Jackson 2011), and culturally appropriate strategies for coping with stress (Abdou et al. 2010). For example, studies show that, regardless of level of income, Black Americans are more likely than Whites to live in stressful environments indexed by segregated neighborhoods (Williams and Collins 2001) and environments such as food deserts that do not promote health (Powell et al. 2007).

Stress and stress-induced eating have received substantial attention in the obesity literature (Tamashiro, Hegeman, and Sakai 2006; Torres and Nowson 2007). Stress can promote obesity by affecting food intake and influencing physiological pathways. In animal models, stress promotes preferential intake of comfort food and highly dense calories, and this in turn appears to blunt the stress response (Dallman 2010), suggesting a functional positive feedback loop that might further induce overeating under stress. In humans, stress also predicts greater intake of comfort food (Epel et al. 2001), and evidence has suggested that the stress response process in animals is similar in humans (Jackson, Knight, and Rafferty 2010). Indeed, studies show that higher stress is associated with lower self-rated diet quality (Burdette and Hill 2008), lower fruit and vegetable consumption (Cartwright et al. 2003), higher intake of fatty foods (Cartwright et al. 2003; Laitinen, Ek, and Sovio 2002), and higher levels of unhealthy snacking (Cartwright et al. 2003).

Another influence of stress is the disruption of optimal lifestyle behaviors that are important for managing stress, such as engaging in positive social interactions and regular moderate to vigorous physical activity, eating a healthy diet, consuming alcohol in moderation, avoiding cigarettes, and higher sleep duration and quality (Sun et al. 2012). Studies show that high levels of perceived stress is associated with smoking (McEwen, West, and McRobbie 2008), alcohol abuse (Tavolacci et al. 2013), and less
frequent physical activity (Stults-Kolehmainen and Rajita 2014), which increase the risk for conditions such as obesity and depression.

Stress-coping strategies in the form of healthy behaviors, such as physical activity, have been shown to mitigate the effects of stress on both mental and physical health outcomes (Craike, Coleman, and MacMahon 2010; Puterman et al. 2010). However, few studies examine the role of stress in the relationship between physical activity and obesity risk among Blacks. Findings from studies that examine the extent to which health behaviors moderate the association between stress and physical and mental health outcomes suggest that poor health behaviors such as poor diet and low physical activity buffer the effects of life stress on depression for Blacks but not for Whites, but increase the risk for chronic health conditions for both groups (Jackson, Knight, and Rafferty 2010; Mezuk and Rafferty 2010). Additional studies are needed to document the effects of low and high physical activity for Black women, especially in light of studies reporting few benefits of physical activity for reducing the obesity risk in this population (West et al. 2008; White and Jago 2012).

**Focus of the present study**

The current study aims to further elucidate the association between physical activity, obesity, and depression among women by examining social stressors associated with both conditions. Specifically, I hypothesize that health-promoting behaviors, such as physical activity, will be associated with lower risk for obesity and depression in African American, Caribbean Black and White women. I also hypothesize that physical activity will moderate the association between stress and obesity, and between stress and depression, thereby clarifying the role of physical activity on obesity and depression risk among women. This study advances research in several ways. The focus on social stress as a correlate to obesity contributes to the broader obesity literature that has primarily focused on health behaviors. The inclusion of depression as an outcome allows for an additional test of the association between stress, poor health behaviors and positive mental health outcomes previously reported by investigators. Further, an examination of physical activity as a moderator addresses studies suggesting that Black females do not benefit from physical activity to the same degree as White females. The use of a large, nationally representative sample of African Americans and Caribbean Blacks from across the United States permits generalizability of study findings and an investigation of potential ethnic group differences in the relationships among study variables. The combined advantages of the sample and study variables contribute to knowledge of the role of social stress in the relationship between physical activity, obesity, and depression, particularly among diverse groups of Black American women.

**Methods**

**Sample**

The National Survey of American Life: Coping with Stress in the 21st Century (NSAL) was collected by the Program for Research on Black Americans at the University of Michigan’s Institute for Social Research. The fieldwork for the study was conducted from 2001 to 2003 and was completed by the Institute of Social Research’s Survey Research Center,
in cooperation with the Program for Research on Black Americans. This study was approved by the University of Michigan’s Institutional Review Board.

Interviews were conducted with 6082 individuals aged 18 or older, including 3570 African Americans, 891 non-Hispanic Whites, and 1621 Blacks of Caribbean descent. The NSAL included the first major probability sample of Caribbean Blacks ever conducted. For the purposes of this study, Caribbean Blacks are defined as individuals who trace their ethnic heritage to a Caribbean country but now reside in the United States, are racially classified as Black, and speak English.

The NSAL used a national multistage probability design. The interviews were conducted face-to-face in respondents’ homes. Respondents were compensated $50 for their time. The overall response rate was 72.3%. Final response rates for the NSAL two-phase sample designs were computed using American Association of Public Opinion Research (American Association for Public Opinion Research 2006) guidelines for Response Rate 3.

Measures

Body mass index (BMI) was calculated by dividing self-reported weight in kilograms by height in meters squared. Respondents were classified as either obese (BMI ≥ 30 kg/m²), overweight (BMI 25–29.9 kg/m²), or not overweight (BMI ≤ 24.9 kg/m²).

Depressive symptoms were assessed using the 12-item version of the Center for Epidemiological Studies-Depression scale (Radloff 1977). This abbreviated CES-D has been found to have acceptable reliability and a similar factor structure compared to the original 20-item version. The 20-item CES-D scale is widely used, with scores ranging from 0 to 60. Similar to the 20-item version, the cutoff score for the 12-item CES-D scale is typically 16 or more and is also considered indicative of potential depression and the need for additional screening (Nguyen et al. 2004). These 12 items measured the extent to which respondents experienced things such as: had trouble keeping their mind on tasks, enjoyed life, had crying spells, felt depressed, hopeful, happy, and that everything was an effort during the previous 30 days. Item responses were coded from 1 (hardly ever) to 3 (most of the time). Positive valence items were reverse coded and summed, resulting in a continuous measure; a high score indicated more depressive symptoms (M = 6.51, SE = 0.23; Cronbach’s alpha = .75).

Physical activity was measured with three questions assessing the frequency with which respondents engaged in sports or physical activity, walking, and gardening or yard work, with a 4-point response scale (1 = never, 2 = rarely, 3 = sometimes, and 4 = often). Higher scores indicated more physical activity.

To assess chronic stress, respondents reported whether they had experienced any of 10 chronic stressors during the previous month. Sample stressors included problems related to health, money, employment, law enforcement, or family. Affirmative answers were summed to create a continuous measure, with higher scores reflecting more stressors experienced during the previous month.

Material hardship was assessed by asking seven questions to ascertain whether or not certain events occurred. For example, respondents were asked, ‘In the past 12 months was there a time when you …’ (1) ‘didn’t pay full rent or mortgage’; (2) ‘were evicted for non-payment’; (3) ‘had gas, electric, oil disconnected’; and (4) ‘had the telephone
disconnected?’ Affirmative answers were summed and higher scores reflect greater material hardship.

Experiences of unfair treatment were assessed using the Everyday Discrimination Scale, consisting of 10 items designed to measure routine experiences of unfair treatment (Williams et al. 1997). This measure of unfair treatment has been widely used in studies of discrimination and health (McIlvane, Baker, and Mingo 2008). Items were measured on a 6-point scale ranging from 0 (never) to 5 (almost every day). Respondents were asked if they had experienced being treated with less courtesy or less respect; receiving poorer service; perceiving that other people were acting afraid of them or as if they are better than them; being perceived as less smart or dishonest; and being insulted, threatened, and followed around in stores due to their race. Total scores ranged from 0 to 50, with higher scores reflecting more experiences of unfair treatment.

The demographic variables used in this analysis included race and ethnicity, age, education, and family income. Missing data for family income and education were imputed using an iterative regression-based multiple imputation approach incorporating information about age, sex, region, race, employment status, marital status, home ownership, and nativity of household residents. All analyses used the SVY survey analysis procedures of STATA version 11.2, which provides estimates that account for the incorporation of complex survey sampling methods including multistage and cluster study designs (StataCorp 2007). Unweighted frequencies were used to describe characteristics of the NSAL data (Table 1).

### Table 1. Characteristics of the sample.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>African American n (%)</th>
<th>Caribbean Black n (%)</th>
<th>Non-Hispanic White n (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–29</td>
<td>480 (25.00)</td>
<td>217 (27.12)</td>
<td>56 (24.33)</td>
<td>0.692</td>
</tr>
<tr>
<td>30–44</td>
<td>745 (35.00)</td>
<td>340 (37.96)</td>
<td>86 (32.18)</td>
<td></td>
</tr>
<tr>
<td>45–59</td>
<td>463 (23.52)</td>
<td>195 (18.84)</td>
<td>99 (23.26)</td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>344 (16.48)</td>
<td>112 (16.08)</td>
<td>97 (20.23)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>≤11 years</td>
<td>525 (24.98)</td>
<td>144 (20.47)</td>
<td>53 (13.39)</td>
<td></td>
</tr>
<tr>
<td>12 years</td>
<td>759 (36.26)</td>
<td>262 (30.22)</td>
<td>104 (30.53)</td>
<td></td>
</tr>
<tr>
<td>13–15 years</td>
<td>468 (24.65)</td>
<td>239 (29.37)</td>
<td>81 (23.77)</td>
<td></td>
</tr>
<tr>
<td>16+ years</td>
<td>280 (14.11)</td>
<td>219 (19.94)</td>
<td>100 (32.31)</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>&lt;$25,000</td>
<td>1160 (51.23)</td>
<td>340 (39.45)</td>
<td>126 (32.03)</td>
<td></td>
</tr>
<tr>
<td>$25,000–$49,999</td>
<td>544 (27.85)</td>
<td>295 (32.09)</td>
<td>111 (33.38)</td>
<td></td>
</tr>
<tr>
<td>$50,000–$74,999</td>
<td>200 (11.98)</td>
<td>132 (16.45)</td>
<td>52 (18.81)</td>
<td></td>
</tr>
<tr>
<td>$75,000–$99,999</td>
<td>84 (6.06)</td>
<td>50 (3.97)</td>
<td>24 (7.96)</td>
<td></td>
</tr>
<tr>
<td>$100,000–$124,999</td>
<td>26 (1.83)</td>
<td>20 (3.39)</td>
<td>11 (3.40)</td>
<td></td>
</tr>
<tr>
<td>$125,000+</td>
<td>18 (1.04)</td>
<td>27 (4.65)</td>
<td>14 (4.42)</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>Underweight</td>
<td>38 (2.33)</td>
<td>12 (1.45)</td>
<td>2 (0.24)</td>
<td></td>
</tr>
<tr>
<td>Healthy weight</td>
<td>556 (27.17)</td>
<td>280 (36.57)</td>
<td>142 (46.53)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>587 (29.61)</td>
<td>311 (30.86)</td>
<td>82 (21.32)</td>
<td></td>
</tr>
<tr>
<td>Obese class 1</td>
<td>461 (21.96)</td>
<td>156 (14.90)</td>
<td>60 (18.33)</td>
<td></td>
</tr>
<tr>
<td>Obese class 2</td>
<td>229 (11.00)</td>
<td>60 (9.68)</td>
<td>36 (8.88)</td>
<td></td>
</tr>
<tr>
<td>Obese class 3</td>
<td>161 (7.93)</td>
<td>45 (6.53)</td>
<td>16 (4.71)</td>
<td></td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td></td>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>0</td>
<td>273 (13.21)</td>
<td>125 (17.91)</td>
<td>16 (5.33)</td>
<td></td>
</tr>
<tr>
<td>1–10</td>
<td>1250 (61.47)</td>
<td>564 (64.10)</td>
<td>206 (62.76)</td>
<td></td>
</tr>
<tr>
<td>11–15</td>
<td>291 (14.08)</td>
<td>107 (10.38)</td>
<td>69 (18.91)</td>
<td></td>
</tr>
<tr>
<td>16 or more</td>
<td>218 (11.23)</td>
<td>68 (7.62)</td>
<td>47 (13.01)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Unweighted frequencies, weighted percentages.
**Results**

Characteristics of the sample are presented in Table 1. There were no significant age differences between African American, Caribbean Black, and White women. On average, African American women had lower levels of education and income and higher BMI than Caribbean Black and White women. African American and Caribbean Black women reported fewer depressive symptoms compared to White women.

Multivariable regression analysis of the effects of chronic stress, material hardship, and racial discrimination on obesity, controlling for demographic variables is presented in Table 2. Age was positively associated with BMI for African American women across all three models (e.g. chronic stress, material hardship, and racial discrimination). Specifically, older age was associated with higher risk for obesity for African American women. Education was positively associated with BMI among African American women such that higher levels of education was associated with higher risk for obesity.

Age was negatively associated with depressive symptoms for African American and Black Caribbean women across all three models; those of advanced age had a lower risk for depression. Education was negatively associated with depressive symptoms for African American and Black Caribbean women such that higher levels of education was associated with lower risk of depression. Finally, income was negatively associated with lower risk of depression for African American and Black Caribbean women; higher income was associated with fewer depressive symptoms.

Since this study is a new area of inquiry, the effects of each social stressor were estimated in separate models to determine their independent and multiplicative effects with physical activity on obesity and depressive symptoms. Among African American and White women, stress was not related to BMI. Among Caribbean Black women, however, chronic stress was associated with higher BMI. Physical activity was associated with lower BMI for all women across all three models.

The multivariable analysis of the effects of chronic stress, material hardship, and racial discrimination on depressive symptoms, controlling for all demographic variables is presented in Table 3. Results indicated that chronic stress, material hardship, and discrimination were all associated with higher levels of depressive symptoms for African American and White women after controlling for other covariates. Among Caribbean Black women, only chronic stress was associated with more depressive symptoms. Physical activity was associated with fewer depressive symptoms for African American and White women across all three models. For Caribbean Black women, physical activity was only associated with fewer depressive symptoms in the model containing discrimination.

**Interaction models**

The hypothesis that physical activity moderates the effect of stress on BMI and depressive symptoms was tested. Physical activity moderated the effect of chronic stress on BMI (Figure 1; Table 4) and depressive symptoms BMI (Figure 2; Table 5) for African American women reporting high levels of chronic stress but not for those reporting low levels of chronic stress. Physical activity did not appear to have the same benefit for women who experienced low levels of chronic stress. Among White women, physical activity
Table 2. Weighted linear regressions of obesity among African American, Caribbean Black, and non-Hispanic white women in the National Survey of American Life.

<table>
<thead>
<tr>
<th></th>
<th>Chronic stress</th>
<th>Material hardship</th>
<th>Racial discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>African American</td>
<td>Black Caribbean</td>
<td>White</td>
</tr>
<tr>
<td>Age</td>
<td>0.032*</td>
<td>0.001</td>
<td>0.008</td>
</tr>
<tr>
<td>Education</td>
<td>−0.223***</td>
<td>−0.263</td>
<td>−0.285</td>
</tr>
<tr>
<td>Income</td>
<td>−0.006</td>
<td>0.027</td>
<td>0.013</td>
</tr>
<tr>
<td>Chronic stress</td>
<td>1.003</td>
<td>5.457***</td>
<td>1.585</td>
</tr>
<tr>
<td>Material hardship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td>−0.615**</td>
<td>−2.264***</td>
<td>−1.868**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.

Table 3. Weighted linear regressions of depressive symptoms among African American, Caribbean Black, and non-Hispanic white women in the National Survey of American Life.

<table>
<thead>
<tr>
<th></th>
<th>Chronic stress</th>
<th>Material hardship</th>
<th>Racial discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>African American</td>
<td>Black Caribbean</td>
<td>White</td>
</tr>
<tr>
<td>Age</td>
<td>−0.047***</td>
<td>−0.033**</td>
<td>0.007</td>
</tr>
<tr>
<td>Education</td>
<td>−0.446***</td>
<td>−0.366***</td>
<td>−0.147</td>
</tr>
<tr>
<td>Income</td>
<td>−0.013*</td>
<td>−0.010*</td>
<td>−0.008</td>
</tr>
<tr>
<td>Chronic stress</td>
<td>8.768***</td>
<td>8.425***</td>
<td>7.990***</td>
</tr>
<tr>
<td>Material hardship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td>−0.557***</td>
<td>−1.053</td>
<td>−1.936*</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.
moderated the effect of racial discrimination on BMI for women who reported experiencing both high and low levels of discrimination (Figure 3). White women who reported fewer experiences of discrimination had a higher BMI compared to those who reported more experiences of discrimination. However, physical activity had a more robust effect on the BMI of women who reported fewer experiences of discrimination and as a result, their BMI was lower than that of women who reported more discrimination. Physical activity was not a significant stress moderator for Caribbean Black women in the models of depressive symptoms or BMI.
Table 4. Weighted linear regressions of depressive symptoms including the interaction between stress and physical activity among African Americans, Caribbean Blacks and non-Hispanic whites, National Survey of American Life.

<table>
<thead>
<tr>
<th></th>
<th>Chronic stress</th>
<th>Material hardship</th>
<th>Racial discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>African American</td>
<td>Black Caribbean</td>
<td>White</td>
</tr>
<tr>
<td>Age</td>
<td>−0.046***</td>
<td>−0.033*</td>
<td>0.007</td>
</tr>
<tr>
<td>Education</td>
<td>−0.439***</td>
<td>−0.372***</td>
<td>−0.149</td>
</tr>
<tr>
<td>Income</td>
<td>−0.014**</td>
<td>−0.010*</td>
<td>−0.008</td>
</tr>
<tr>
<td>Chronic stress</td>
<td>13.711***</td>
<td>7.104</td>
<td>12.637</td>
</tr>
<tr>
<td>Material hardship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td>−0.017</td>
<td>−1.177**</td>
<td>−1.572</td>
</tr>
<tr>
<td>Stress × physical activity</td>
<td>−1.948***</td>
<td>0.505</td>
<td>−1.700</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.

Table 5. Weighted linear regressions of obesity including the interaction between stress and physical activity among African Americans, Caribbean Blacks and non-Hispanic whites, National Survey of American Life.

<table>
<thead>
<tr>
<th></th>
<th>Chronic stress</th>
<th>Material hardship</th>
<th>Racial discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>African American</td>
<td>Black Caribbean</td>
<td>White</td>
</tr>
<tr>
<td>Age</td>
<td>0.033*</td>
<td>0.001</td>
<td>0.008</td>
</tr>
<tr>
<td>Education</td>
<td>−0.217**</td>
<td>−0.252</td>
<td>−0.283</td>
</tr>
<tr>
<td>Income</td>
<td>−0.007</td>
<td>0.027</td>
<td>0.013</td>
</tr>
<tr>
<td>Chronic stress</td>
<td>5.457*</td>
<td>8.093</td>
<td>−4.351</td>
</tr>
<tr>
<td>Material hardship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td>−0.129</td>
<td>−2.017**</td>
<td>−3.344</td>
</tr>
<tr>
<td>Stress × physical activity</td>
<td>−1.755*</td>
<td>−1.007</td>
<td>2.171</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.
Discussion

The current study contributed to the literature examining the role of physical activity in reducing the risk of obesity and depression among women by considering the role of social stress in these relationships. Furthermore, findings extend the existing literature by providing preliminary evidence that the benefit of physical activity for African American women is dependent on the type and level of stress exposure.

Results of the direct-effects models indicated that chronic stress was associated with higher BMI for Caribbean Black women and that chronic stress, material hardship, and discrimination were associated with more depressive symptoms for African American and White women. These findings highlight an important ethnic distinction between African Americans and Caribbean Blacks. To date, few studies have examined ethnic differences in the Black American population. Consequently, explanations for the differing effects of chronic stress observed between these two groups are largely speculative. However, previous studies have noted the cultural differences between African Americans and Caribbean Blacks due to immigration status, language, religion, traditions, and family structure. These important distinctions have been attributed to observed differences between Caribbean Blacks and African Americans in outcomes such as obesity (Lincoln, Abdou, and Lloyd 2014), social support (Lincoln, Taylor, and Chatters 2013), stress appraisals (Goosby et al. 2012), and religious practices (Chatters et al. 2009). Although Black Caribbeans and African Americans share similar characteristics in the United States, including living in the same neighborhoods and experiencing discrimination (Waters 1997), their immigrant origins and cultures may contribute to differences in manifestations of stress and coping (Georgiades, Boyle, and Duku 2007). Thus, differences in stress appraisals and coping strategies among African American versus Caribbean Black women may explain differential findings regarding the effects of chronic stress and demonstrate the importance of considering an array of social stressors to better capture
heterogeneity within and between populations, including different levels and types of stress exposure, vulnerability, and experiences.

Physical activity was associated with lower BMI for all three groups of women across all three models (i.e. chronic stress, material hardship, discrimination), and with fewer depressive symptoms for African American and White women. These findings are consistent with those of previous cross-sectional (Dunn, Trivedi, and O’Neal 2001) and longitudinal studies that reported positive benefits of physical activity for obesity (Centers for Disease Control and Prevention 2011) and depressive symptoms (Mikkelsen et al. 2010). Again, findings also highlight the importance of considering ethnicity in the Black American population, because the findings for Caribbean Black women diverged significantly from those of African American women.

Findings from regression analysis examining physical activity as a moderator revealed interesting points of divergence between African American and White women that were not evident in the direct-effects models. Findings indicated that physical activity moderated the effects of stress on obesity and depression, but primarily for African American women. Specifically, physical activity did not further reduce the risk of depression for African American women who reported experiencing low levels of chronic stress. However, physical activity significantly reduced the effect of chronic stress on depressive symptoms for women who reported experiencing high levels of chronic stress.

These findings are consistent with the stress-buffering hypothesis (Cohen and McKay 1984), which posits that a coping resource used in response to stress will be more beneficial for individuals experiencing high levels of stress than for those experiencing no or low levels of stress. In the context of the current study, physical activity was beneficial for all women; however, it was more beneficial for African American women who experienced high levels of chronic stress. This finding is supported by other studies examining the stress-buffering potential of physical activity. One study testing the stress-buffering function of physical activity on accelerated cellular aging among postmenopausal women found that vigorous physical activity (compared to being sedentary) protected those experiencing high stress by buffering its relationship with telomere length, (Puterman et al. 2010) a biomarker for biological age that has been linked to health problems including coronary heart disease, diabetes, and early mortality (Fitzpatrick et al. 2006). Another study found that physical activity was especially helpful for reducing the effects of major life stressors on depression. Essentially, a person who had experienced one negative life event and who was very physically active was expected to have approximately the same global depression score as a sedentary person with no negative life events.

Among White women, physical activity moderated the effect of racial discrimination on BMI for women who reported experiencing both high and low levels of discrimination. The parameter estimates should be interpreted with caution due to the relatively small number of White participants reporting racial discrimination (6.2% vs. 50.5% for African Americans). Nonetheless, the strength of the association ($p < .01$) suggests that the finding is important and worthy of discussion. Although the majority of studies of racial discrimination and obesity have largely focused on Blacks, perceived unfair treatment due to race or ethnicity has been shown to be adversely related to the health of multiple racial and ethnic population groups in the United States (Ryan, Gee, and Laflamme 2006) and internationally (De Vogli et al. 2007; Harris et al. 2006). Hunte and Williams (Hunte and Williams 2009) examined the relationship between perceived discrimination and obesity in a multiracial and
multiethnic sample of adults and found that perceived experiences of unfair treatment were associated with increased abdominal obesity, but only among Whites. This finding was not evident among Hispanics or African Americans.

In light of this finding, it is important to point out that the non-Hispanic White respondents in this sample are unique because they live in the same contexts and geographical areas as African Americans are distributed (both high and low density, urban and rural, inner-city and suburban, etc.). Although they reside in similar contexts and in relatively close proximity to African Americans, the White sample had higher levels of education, income, and employment rates than African Americans, and were more likely to own their homes (Jackson et al. 2004). It is also important to note that African Americans experience racial discrimination more frequently and with greater intensity than Whites (Kessler, Mickelson, and Williams 1999; LaVeist, Rolley, and Diala 2003; Williams et al. 1997). Moreover, Black women experience a unique form of oppression that is specific to their race and their gender (Thomas, Witherspoon, and Speight 2008).

With these points in mind, perhaps the adverse reaction to perceived racial discrimination among White women is due to less developed coping responses compared to African American women, who have historically experienced racial discrimination across a variety of contexts. It is also possible that Whites who live in neighborhoods where they are not the majority might experience discrimination from a different racial group (intergroup) or from members of the same racial group (intragroup discrimination). That is, perceived discrimination among Whites might arise in the face of heightened interracial contact where interactions that reflect class disadvantages may also be understood as racially charged.

**Conclusion**

This study highlights the importance of considering social stress to understand the role of physical activity in reducing the risk for obesity and depression in women. Findings for African American women, in particular, provide some insight into studies that reported little or no benefit of physical activity for Black females (White and Jago 2012; West et al. 2008). The current study contributes to the idea that racial disparities in obesity are not solely due to biological or genetic differences or health behaviors. Stress exposure, stress appraisals, and the overall social context in which individuals live, should also be considered.

The finding that physical activity moderated the effects of high levels of chronic stress on BMI and depression for African American women suggests that weight loss and weight management interventions should also include strategies for managing stress along with promoting health behavior change. Such strategies are particularly challenging for those residing in neighborhoods that are overly hazardous (e.g. pollution, crime) and with limited access to healthy foods and outlets for physical activity (e.g. parks, walking trails). Current weight-loss interventions that focus on behavior change have had limited success, particularly for African American women (West et al. 2008). Clearly, the role of social stress and the built environment should be considered in future interventions.

Findings from this study should be considered within the context of its strengths and limitations. First, because several segments of the population were not represented, such as men and institutionalized individuals, findings are not generalizable to these subgroups. Second, causal inferences are problematic with cross-sectional data because we have no
information about the onset of the outcome statuses, and longitudinal data are preferred for the interpretation of social status effects (e.g. education, income). Third, height and weight were assessed by self-report. Although this is a common approach (Kuczmarski, Kuczmarski, and Najjar 2001) and the biases in these types of self-reported data are fairly well understood, the bias associated with underestimating weight and overestimating height should be noted (Barry, Pietrzak, and Petry 2008). Fourth, social stress measures were also assessed by self-report. Although this concern is likely widely shared, the use of self-reported measures of stress is quite common (Masood et al. 2012). In addition, studies have shown that perceived measures of stress are indeed more accurate and relevant for assessing health outcomes compared to objective measures (Masood et al. 2012). Fifth, the measure of physical activity reflected only a small portion of potential physical activities and did not capture information about the duration or intensity of such activities. If more comprehensive and precise measures were available, the observed effects may have been stronger and other relationships may have emerged. Finally, separate analysis was not conducted for different ethnic groups in the Black Caribbean population due to small sample sizes for some ethnic groups. Although the homogenization of the Black Caribbean population is not ideal, this study had substantial strength to distinguish between African Americans and Black Caribbeans as an initial step in understanding within-group ethnic differences for the Black population using nationally representative data. Given that this study revealed important intraracial differences in the manifestation of stress in African American and Black Caribbean women, more systematic studies are needed to understand the potential ethnic differences in coping strategies employed by these groups. For researchers and policy makers to effectively alleviate depression and obesity risk in the Black population, more rigorous studies acknowledging and examining variation in the stress process across ethnic groups are needed.

Despite these limitations, this study was the first to document the moderating role of physical activity in the relationship between stress, depression, and obesity using a nationally representative sample of African American, Caribbean Black, and White women. Findings provide insight into the role of stress in relation to physical activity, depression, and obesity while also highlighting heterogeneity among Black American women. Findings also offer important prescriptions for the design of interventions targeting weight loss and management for Black women.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Southern California Clinical and Translational Science Institute, Keck School of Medicine, University of Southern California [grant number UL1 RR031986-02].

Key messages

(1) Physical activity is beneficial for all women; however, it was more beneficial for African American women who experienced high levels of chronic stress.
Results indicating differences between African Americans and Caribbean Blacks in the association between stress, obesity and depression highlight the importance of considering ethnic distinction between these two groups.

Racial disparities in obesity are not solely due to biological or genetic differences or health behaviors. Stress exposure, stress appraisals, and the overall social context in which individuals live, should also be considered.

References


StataCorp. 2007. STATA: Release 10.0. College Station, TX: StataCorp LP.


Race-related Differences in Depression Onset and Recovery in Older Persons Over Time: The Health, Aging, and Body Composition Study

Lisa C. Barry,
Center on Aging, University of Connecticut, Farmington, Connecticut 06030-5215, US

Roland J. Thorpe Jr.,
Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, US

Brenda W.J.H. Penninx,
Department of Psychiatry, VU University Medical Center in Amsterdam, The Netherlands

Kristine Yaffe,
Department of Psychiatry, University of California, San Francisco, California, US

Dorothy Wakefield,
Center on Aging, University of Connecticut, Farmington, Connecticut, US

Hilsa N. Ayonayon,
Department of Epidemiology and Biostatistics, University of California, San Francisco, California, US

Suzanne Satterfield,
Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, US

Anne B. Newman, and
Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, US

Eleanor M. Simonsick
Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, US

Abstract
**Objectives**—To evaluate race-related differences in depression onset and recovery in older persons, overall and by sex, and examine race-related differences in mortality according to depression.

**Design**—Prospective cohort study.

**Setting**—General community in pre-designated zip code areas in Memphis, Tennessee, and Pittsburgh, Pennsylvania.

**Participants**—3,075 persons age 70-79 at baseline in the Health, Aging, and Body Composition study.

**Measurements**—Depression was assessed at 8 time points over 10 years using the 10-item Center for Epidemiologic Studies – Depression (CES-D) scale; scores of <8 and ≥8 denoted nondepressed and depressed, respectively. We created variables for transitions across each 18-month time interval, namely, from nondepressed or depressed to nondepressed, depressed, or death, and determined the association between race and the average likelihood of these transitions over time.

**Results**—A higher percentage of blacks than whites were depressed at nearly all time points. Adjusting for demographics, common chronic conditions, and body mass index, blacks had a higher likelihood of experiencing depression onset than whites (odds ratio, 1.22; 95% confidence interval, 1.03-1.43); among men, blacks were more likely to experience depression onset than whites (odds ratio, 1.44; 95% confidence interval, 1.24-2.89). Blacks also had a higher likelihood of transitioning from nondepressed to death (odds ratio, 1.79; 95% confidence interval, 1.30-2.46). Overall and in sex-stratified analyses, race was not associated with recovery from depression or with the transition from depression to death.

**Conclusion**—Our findings highlight race differences in depression in older persons and encourage further research on the course of depression in older blacks.

**Keywords**
Aging; Depression; Depressive symptoms; Race differences; Epidemiology; Prospective studies

**OBJECTIVE**

Depression is a significant clinical and public health problem for older persons. Whereas major depression affects 1%–2% of community-living persons age 65 and older (1), 8%–30% of this population experiences clinically significant depressive symptoms (1, 2). Clinically significant depressive symptoms, often referred to simply as “depression,” are associated with negative outcomes including high health services utilization, disability, dementia, and mortality (3-8). What we know about the course of depression in older persons, however, has come largely from overwhelmingly white samples (9-11).

Older Americans are becoming increasingly diverse, with blacks projected to account for at least 12% of persons ≥65 living in the U.S. by 2050 (12). Despite growing numbers of blacks living to older ages, and the national public health goal of reducing health disparities between minority and non-minority populations (13), whether or not race differentially influences the course of depression in older persons remains uncertain.
Depression is highly variable over time, with periods of chronicity, remission and recurrence (14, 15). Yet, most studies evaluating race-related differences in depression have been cross-sectional, focusing on prevalence estimates or number of symptoms at one time point as outcomes (16-19). Furthermore, the few longitudinal studies evaluating race-related differences in the course of depression in older persons have focused on symptom trajectories and average symptom levels rather than transitions into and out of clinically meaningful states of depressive symptoms (20-23). Of these, only two considered the potential impact of mortality on race differences in depression over time (20, 23). Additionally, the same non-modifiable factors that predict depression onset may not predict the course of depression. For example, low educational attainment and female sex have been found to be associated with depression onset (24-26) but not remission (27). Relatedly, female sex has been found to be associated with more severe depressive symptoms (28) and a higher likelihood of receiving treatment for depression (29), while black race is associated with a lower likelihood of initiation and adherence to depression treatment(30). Consequently, whether there are race-related differences in the onset of and recovery from depression in older persons, or whether race is similarly associated with these outcomes in men and women, remains unknown.

The objectives of this prospective study were to determine whether there are race-related differences, overall and by sex, in the onset of depression (i.e., first or recurrent episodes) and recovery from depression in older persons over time and to examine race-related differences in mortality according to depression. To achieve these objectives, we evaluated possible transitions between three states—not depressed, depressed, and death—over 10 years in a large cohort of older men and women. Improved understanding of these race-related differences will support efforts to identify and subsequently reduce racial disparities in depression treatment and prevention.

METHODS

Study Population

Participants were members of the Health, Aging, and Body Composition (Health ABC) study. Assembly of the cohort (N=3,075) has been described elsewhere(31). In brief, participants constitute a random sample of white Medicare beneficiaries and all age-eligible black community-dwelling residents in pre-designated zip code areas in Memphis, Tennessee, and Pittsburgh, PA, who were aged 70 to 79 years at time of enrollment between April 1997 and June 1998. At baseline, all participants reported no difficulty walking 0.25 miles, walking up ten steps without resting, and performing basic activities of daily living. Participants signed a written informed consent and protocols were approved by the study site institutional review boards.

Data Collection

Health ABC data were collected using both face-to-face interviews and clinic visits that occurred over 10 years of follow-up. Because depressive symptoms were not assessed during years 2, 7 and 9, data from years 1, 3, 4, 5, 6, 8, 10, and 11 were included in the present study. Participants unable to attend their clinic visit were offered a home visit or a
telephone follow-up. Deaths were ascertained by proxy report and were confirmed by obtaining death certificates. A total of 1,160 (37.7%) participants died after a mean (SD) follow-up of 5.7 (±2.6) years.

Demographic data, including race, age, sex, years of education, and poverty status were collected during the Year 1 (i.e., baseline) face-to-face interview. Participants self-identified their race as either white or black. Education was categorized as less than high school and high school or more. Poverty was determined by applying a set of income thresholds that vary by family size and composition issued in 1998 by the U.S. Census Bureau to reported annual income levels. Every participant was assigned a value representing a percentage of the federal poverty level. Participants with an income level at or below 150% of the federal poverty level were considered to have low socioeconomic status (SES). This indicator, commonly used in determining eligibility for federal and state public assistance programs, has been used previously with these data (32). Presence of coronary heart disease (CHD), stroke, diabetes, and cancer were ascertained at baseline and during subsequent face-to-face interviews by asking participants if a doctor told them that they had each of these conditions. Body mass index (BMI), reported as kg/m², was calculated from measured weight and height collected during the clinic assessments.

Assessment of depressive symptoms and transition states

Depressive symptoms were assessed during the face-to-face interviews at years 1, 3, 4, 5, 6, 8, 10, and 11 using the 10-item Center for Epidemiologic Studies – Depression (CES-D) scale. Participants were asked about the frequency of depressive symptoms in the past week. Total scores range from 0 to 30, and participants scoring ≥8 were considered as “depressed” (33). Data on depression were complete for 99% of the participants at baseline and 96%, 94%, 94%, 76%, 78%, and 80% of the 3075 participants at years 3, 4, 5, 6, 8, 10, and 11, respectively. Because the CES-D was not administered to study participants at all visit types in years 8, 10 (i.e., home visit, telephone interview), and 11 (e.g., home visit), the total percent with depression data was lower than in previous years. However, of the participants who could have completed the CES-D in years 8, 10, and 11, less than 1% were missing depression data.

We then created variables for six possible transitions that could occur across each of the 18-month time intervals. The transitions were nondepressed to nondepressed, nondepressed to depressed (i.e., depression onset), or nondepressed to death; and depressed to nondepressed (i.e., recovery), depressed to depressed; or depressed to death. We have used this approach previously to evaluate transitions into and out of depression states over time (24). Because there were seven time intervals over the 10-year study period (e.g., years 1 to 3; years 3 to 4), each participant had an opportunity to contribute up to a total of seven transitions.

Statistical Analyses—We compared baseline participant characteristics by race using χ² or t test statistics. We then used the χ² statistic to determine if the percentage of whites and blacks with depression at each time point was significantly different. To determine whether the observed transitions into and out of depression were clinically meaningful, we calculated the percentage of transitions that represented absolute changes in the CES-D scores ≥4
points for each of the 18-month time intervals. We subsequently determined whether the distribution of these percentages differed by race using the χ² statistic.

Given the likely intracorrelation between estimates over time, running a polytomous logistic regression model at each time interval would likely yield biased results. Consequently, we used longitudinal statistical methods to determine if, on average over time, race was associated with depression onset (i.e., transition from a nondepressed to a depressed state) and recovery from depression (i.e., transition from a depressed to a nondepressed state) and to examine race-related differences in the average likelihood of transitioning from a nondepressed or a depressed state to death. Specifically, we ran generalized multinomial logit models for nominal outcomes that were estimated with a generalized estimating equation (GEE) and used exchangeable correlation structures. We ran one longitudinal model which included all participants who were nondepressed at the beginning of any of the 7 time intervals. Participants who remained nondepressed during an interval (i.e., at two consecutive time points) were used as the reference group. Consequently, the parameter estimates for the model reflected the population values averaged across the respective strata; nondepressed to depressed and nondepressed to death versus remaining nondepressed. The second model included participants who were depressed at the beginning of any of the 7 time intervals. Participants who remained depressed during an interval were used as the reference group and the parameter estimates reflected the population values averaged across the respective strata; depressed to nondepressed and depressed to death versus remaining depressed. Final models were adjusted for age, sex, race, education, poverty, chronic conditions, BMI, and study site. We then ran the above models with a Race X Sex interaction term, and subsequently conducted sex-stratified analyses. Because the CES-D was not administered to study participants at all visit types in years 8, 10, and 11, we re-ran the models using only the data from years 1, 3, 4, 5, and 6 for comparison purposes. Finally, to complement the aforementioned analyses, and confirm the findings from prior studies, we repeated these analyses after considering depression as a continuous variable. We used generalized estimating equations to fit multivariate, marginal, linear regression models to determine if the average depression score over time differed by race.

Data were analyzed using SAS version 9.3 (SAS Institute, Inc., Cary, NC). Longitudinal models were performed using SUDAAN statistical software version 11.0 (RTI Institute, Research Triangle Park, NC). The threshold for statistical significance was α ≤0.05 (two-tailed).

RESULTS
Baseline Characteristics

Table 1 provides the baseline characteristics of the sample overall and according to race. The average age of the 3,075 participants was 73.6 (±2.9) years and approximately half were female and half were recruited from each site. Approximately one quarter of the sample did not graduate from high school and 46.2% were at or below the 150% poverty level. Cancer was the most prevalent chronic condition, reported by approximately 18% of the sample, followed by diabetes, CHD, and stroke. The average BMI was 27.4 (±4.8). Differences by race were observed for all characteristics with the exception of CHD. Blacks’ average age
was older and BMI was higher than whites’ average age and BMI, and a higher percentage of blacks were female, were at or below the 150% poverty level, reported having stroke and diabetes, and lived in Pittsburgh. A higher percentage of whites reported having cancer. These race differences were largely observed within men and women, with the exception of age and CHD; average age did not differ by race for men, and among women a higher proportion of blacks had CHD (data not shown).

A total of 310 (10.1%) participants were depressed at baseline (i.e., CES-D ≥ 8), with no difference according to race (chi-square test statistic = 1.69; df=1; P=0.19). There were 19.5%, 18.4%, 22.5%, 23.6%, 23.1%, 23.2%, and 21.9% depressed participants at years 3, 4, 5, 6, 8, 10, and 11, respectively and a significantly higher percentage of blacks than whites were depressed at each time point, with the exception of years 10 and 11 where there was no race difference. Of the nondepressed at baseline, 1,030 (37.6%) experienced depression at some other point during the study. Whereas there were no race-related differences in the percentage of depressed women at any time point, black men were significantly more likely to be depressed than white men at nearly all time points (Figure 1).

The majority of transitions from a nondepressed state to a depressed state, or vice versa, were based on absolute changes in CES-D scores of ≥4 points. Of the participants who transitioned from a nondepressed to a depressed state, 86%; 69%; 77%; 66%; 69%; 66%; and 58% had a ≥4-point absolute change in their CES-D score over each time interval, respectively. Of those who transitioned from a depressed to a nondepressed state, 70%; 74%; 67%; 77%; 72%; and 75% had a ≥4-point absolute change in their CES-D score over each time interval, respectively. There were no differences in the distribution of change scores by race for either transition.

Table 2 presents the results from the longitudinal models evaluating the association between race and the likelihood of each of the six possible transitions over 10 years. Among the nondepressed, blacks were significantly more likely to transition to a depressed state (i.e., onset) and to death. Among the depressed, there were no race-related differences in transition to a nondepressed state (i.e., recovery) or death. The Race X Sex interaction term was statistically significant in the model evaluating the association between race and transitions among those who were nondepressed (Wald chi-square test statistic = 3.87; df=2; P=0.02), but was not significant in the model evaluating the association between race and transitions among those who were depressed (Wald chi-square test statistic = 0.20, df=2; P=0.82). Figure 2 provides the odds ratios and 95% confidence intervals from the sex-stratified models. In comparison to white men, black men had a higher likelihood of depression onset (t statistic=2.99; df=1402; P=0.003) and of transitioning from nondepressed to death (t statistic=2.97; df=1402; P=0.003). In contrast, there were no race-related differences in the likelihood of depression onset (t statistic=0.71; df=1464; P=0.48) or in the transition from a nondepressed state to death (t statistic=1.88; df=1464; P=0.06) among women. In both men and women, there were no race differences in the likelihood of recovery from depression or transitioning from depression to death. After rerunning the models using only the data from years 1, 3, 4, 5, and 6, the results were similar. Finally, results from modeling CES-D as a continuous variable indicate that, on average, blacks’ CES-D scores were 0.50 (SE 0.12) points higher than whites’ scores (Chi-square score test
statistic 15.76; df=1; \( P<0.001 \). Among men, blacks’ average CES-D scores were 0.79 (SE 0.17) points higher than whites’ scores (Chi-square score test statistic 20.2; df=1; \( P<0.001 \)), whereas there was no race difference in average CES-D scores over time among women (Chi-square score test statistic 2.3; df=1; \( P=0.14 \)).

CONCLUSIONS

In this longitudinal study, which included multiple assessments of depressive symptoms in older persons over 10 years, we found that blacks were more likely than whites to experience depression onset over time and to transition from a non-depressed state to death. Yet, race was not associated with recovery from depression, nor was it associated with a transition from a depressed state to death. Sex-stratified analyses revealed that black men may be more likely than white men to transition from a non-depressed state to either a depressed state or to death. Among women, however, race was not associated with depression onset or recovery, nor was it associated with a transition to death from either a nondepressed or depressed state.

It is uncertain if race is differentially associated with depression in older people. Mixed findings regarding the cross-sectional relationship between race and depression contribute to this uncertainty. Whereas major depression has been found to be more prevalent among older whites than blacks (34), some studies report that blacks have a significantly higher number of depressive symptoms than whites, even after controlling for factors such as SES (16, 19). Other studies report the opposite or have found no association (17, 18). At nearly all time points in the present study, we found that a significantly higher percentage of black participants were depressed as compared with white participants. The stability of this finding, over a long period of follow-up, indicates that the burden of experiencing clinically significant depressive symptoms in older persons may be greater for blacks than whites.

The association between race and depression onset offers additional support for a race-related difference in depression in older persons. We found that blacks were 22% more likely to experience depression onset as compared with whites, even after controlling for important confounders such as number of years of education and poverty. Blacks’ increased likelihood of transitioning from a nondepressed to a depressed state may help to explain the previously reported increase in depressive symptoms over time among older blacks (21) and may also help to explain our finding that blacks have higher average CES-D scores over time as compared with whites. In addition, because older blacks, more so than whites, live in areas with lower-quality healthcare (35), neighborhood characteristics may contribute to the race-related difference in depression onset. The moderate to large changes in the CES-D scores regardless of race, however, are indicative of clinically meaningful transitions from a nondepressed to a depressed state.

In contrast, we found that race was not associated with recovery from depression. This finding was somewhat surprising given recent reports that mental health services use and antidepressant use in older persons is lower among blacks than whites (36-38) and that blacks are less likely to initiate or adhere to treatment for mental health conditions (30, 39). Likelihood of recovery, however, is a function of both diagnosis and appropriate treatment.
Consequently, it is possible that racial disparities in recovery from depression may have emerged had we been able to distinguish (and control for) participants who had been diagnosed with depression by a clinician and subsequently received treatment from those who were undiagnosed. The lack of a race difference in recovery from depression also challenges prior research indicating a higher likelihood of persistent depression among blacks (40). Only 10% of the sample in this prior study, however, included blacks age 60 and older. Future research is needed to verify our findings and evaluate if race differentially influences the effect of diagnosis and initiation of treatment on recovery from depression in older persons. Furthermore, given that 50% of the depressed participants in this study were depressed during at least two consecutive time points, there is still a substantial need for effective recognition and prevention of depression in older persons regardless of race.

Prior research consistently reports that older women are more likely to experience depression than men (24, 41, 42). The under-representation of blacks in these studies, however, has largely precluded the ability to evaluate whether race is differentially associated with depression onset and recovery in men and women separately. Our sex-specific analyses revealed that black men were 44% more likely to experience depression onset than their white counterparts. The difference in depression onset that we observed by sex was marked, with more than a 5-fold difference in the odds ratio for men as compared with women. Relatedly, we also found that average depression scores over time were significantly higher for blacks than for whites among men, yet race was not associated with CES-D score over time among women. Although we controlled for CHD, stroke, diabetes, cancer, and BMI over the study period, it is possible that race differences in other somatic indicators of health, such as kidney disease or tobacco use, which are over-represented in black men (12, 43), may account for the positive association between race and depression onset in men. More research is needed to confirm our finding that black men have a higher likelihood of experiencing depression onset than white men and to identify potentially modifiable factors that may explain this race difference. Additional research may also help to determine the utility of developing preventative efforts that target and are tailored to black men, thereby ultimately supporting the effort to achieve Healthy People 2020's goals of improving mental health through prevention and treatment and reducing health disparities between minority and non-minority populations (13, 44).

By including mortality as an outcome, we addressed the often overlooked issue of competing risk of death in longitudinal studies of older persons (45) that has been found to influence trajectories of depression in a study of adults age 51 and older (20). Blacks in our study were 79% more likely to transition from a nondepressed state to death than were whites. These findings persisted in men, with black men 89% more likely to transition from nondepressed to death as compared with white men. Blacks were also 62% more likely to transition from a depressed state to death than were whites. Although this latter relationship did not reach statistical significance, these substantial effect sizes further reinforce blacks’ mortality disadvantage. The substantial effect sizes among both the nondepressed and depressed groups also warrant additional research to confirm whether or not depression is akin to conditions like flu and pneumonia which confer no race difference in mortality in late life (46).
Use of data from the Health ABC study, a large, racially-diverse study of community-dwelling older adults with excellent retention, provided a unique opportunity to evaluate race differences in depression over a 10-year period. We used longitudinal methods that enhanced statistical power and accounted for potential correlation among depression scores. Furthermore, we not only controlled for important demographic covariates, such as education and poverty status, but we also controlled for time-varying chronic conditions which also have been found to be associated with both race and depression (47, 48).

However, several potential limitations should be acknowledged. All participants were well-functioning at baseline, therefore potentially limiting the generalizability of the findings. We were unable to confirm clinical depression using the CES-D-10. Yet, we used a cut point of ≥8 which has been used previously to indicate clinically significant depressive symptoms (33, 47) that are associated with deleterious health outcomes in older persons (3-6). Because information regarding participants’ depression status before the baseline interview was not available, we could not determine if participants’ first transition from a nondepressed to a depressed state represented incident depression. It is possible that our findings reflect continued race-related patterns of depression established earlier in life. Future studies should evaluate whether race differences in the onset of and recovery from depression during youth and middle age influence depression patterns in older age. Finally, because data on depression was collected at discrete time points, study participants may have experienced additional transitions into and out of depressive states that we could not capture. However, the availability of up to seven waves of follow-up data over 10 years provides the best information, to date, on race-related differences in the onset of and recovery from depression in late life.

Detailed information on the natural history of depression is important to advance efforts in the prevention and treatment of depression in late life. The present research provides new evidence supporting a race difference in the onset of depression, but not recovery, in older persons and further underscores blacks’ mortality disadvantage. We also found evidence to suggest that the relationship between race and depression onset may be stronger in men than in women. Overall, our findings highlight race differences in depression in older persons and reinforce the need for depression prevention strategies. In addition, our findings encourage further research on the course of depression in older blacks, who have historically been under-represented in longitudinal studies of depression in later life.

Acknowledgments

This work was supported in part by the Intramural Research Program of the National Institute on Aging (NIA) contracts N01-AG-6-2101; N01-AG-6-2103; N01-AG-6-2106; the National Institute on Aging at the National Institutes of Health research grants R01-AG028050 and K01-AG-03-1324 (to L.C.B); the National Institute of Nursing research grant R01-NR012459; and National Institute of Mental Health research grant R01-MH086498. The National Institutes of Health provided funding for research on which this article is based.

REFERENCES


Figure 1.
Percentage of depressed (CES-D score >8) White and Black men (Panel A) and White and Black women (Panel B) over 10 years of follow-up, in the Health, Aging, and Body Composition Study. Visit year: year 1, 1997–1998; year 3, 1999-2000; year 4, 2000–2001; year 5, 2001-2002; year 6, 2002–2003; year 8, 2004-2005; year 10, 2007-2008; year 11, 2008-2009. Among men, a significantly higher percentage of Blacks were depressed at years 3, 4, 5, 6, and 11; among women, the percentage of those who were depressed did not differ between Whites at Blacks any year during the study.
Figure 2.
Adjusted odds ratios and 95% confidence intervals derived from sex-stratified multinomial logit models, estimated with Generalized Estimating Equations (GEE), that evaluate the association between race (Blacks versus Whites) and the likelihood of the following transitions: Nondepressed to depressed; Nondepressed to death; Referent group, Nondepressed to nondepressed (odds ratio = 1.0); and Depressed to nondepressed; Depressed to death, Referent group, Depressed to depressed (odds ratio = 1.0).
Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (N=3,075)</th>
<th>White (n=1,794)</th>
<th>Black (n=1,281)</th>
<th>Wald Chi-square or t&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>73.6 (2.9)</td>
<td>73.8 (2.9)</td>
<td>73.4 (2.9)</td>
<td>t=3.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>1584 (51.5)</td>
<td>855 (47.7)</td>
<td>729 (56.9)</td>
<td>25.60</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Not a high school graduate, n (%)</td>
<td>775 (25.2)</td>
<td>219 (12.2)</td>
<td>556 (43.6)</td>
<td>162.08</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>At or below 150% poverty level, n (%)</td>
<td>1421 (46.2)</td>
<td>569 (31.7)</td>
<td>852 (66.5)</td>
<td>405.90</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Coronary Heart Disease, n (%)</td>
<td>357 (11.6)</td>
<td>207 (11.5)</td>
<td>150 (11.7)</td>
<td>0.027</td>
<td>0.87</td>
</tr>
<tr>
<td>Stroke, n (%)</td>
<td>72 (2.3)</td>
<td>25 (1.4)</td>
<td>47 (3.7)</td>
<td>17.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>460 (15.0)</td>
<td>191 (10.7)</td>
<td>269 (21.0)</td>
<td>63.17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>579 (18.8)</td>
<td>439 (24.5)</td>
<td>140 (11.0)</td>
<td>89.37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body Mass Index, mean (SD)</td>
<td>27.4 (4.8)</td>
<td>26.5 (4.1)</td>
<td>28.6 (5.4)</td>
<td>t=−11.38</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pittsburgh, n (%)</td>
<td>1527 (49.7)</td>
<td>859 (47.9)</td>
<td>668 (52.2)</td>
<td>5.44</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<sup>a</sup>Wald chi-square tests have df = 1, t-statistics have df = 3073
Table 2

Longitudinal Assessment of Transitions Over 10 Years Among Blacks Relative to Whites: Health, Aging, and Body Composition Study.

<table>
<thead>
<tr>
<th>Transition</th>
<th>Unadjusted Data</th>
<th></th>
<th></th>
<th>Adjusted Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>95% Confidence Interval</td>
<td>t Statistic</td>
<td>P-value</td>
<td>Odds Ratio</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>Non-depressed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To non-depressed</td>
<td>1.00</td>
<td>Reference</td>
<td>NA</td>
<td>NA</td>
<td>1.00</td>
<td>Reference</td>
</tr>
<tr>
<td>To depressed</td>
<td>1.48</td>
<td>1.30, 1.70</td>
<td>5.74</td>
<td>&lt;0.001</td>
<td>1.22</td>
<td>1.03, 1.43</td>
</tr>
<tr>
<td>To death</td>
<td>1.48</td>
<td>1.27, 1.74</td>
<td>4.86</td>
<td>&lt;0.001</td>
<td>1.79</td>
<td>1.30, 2.46</td>
</tr>
<tr>
<td>Depressed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To depressed</td>
<td>1.00</td>
<td>Reference</td>
<td>NA</td>
<td>NA</td>
<td>1.00</td>
<td>Reference</td>
</tr>
<tr>
<td>To non-depressed</td>
<td>1.07</td>
<td>0.88, 1.30</td>
<td>0.72</td>
<td>0.47</td>
<td>1.14</td>
<td>0.90, 1.45</td>
</tr>
<tr>
<td>To death</td>
<td>1.40</td>
<td>1.03, 1.90</td>
<td>2.15</td>
<td>0.03</td>
<td>1.62</td>
<td>0.89, 2.94</td>
</tr>
</tbody>
</table>

Abbreviations: NA, data Not Applicable.

\(a\) Multinomial logit models that were estimated with Generalized Estimating Equations adjusted for age, sex, education (high school graduate versus did not graduate from high school), socioeconomic status, coronary heart disease, stroke, diabetes, cancer, body mass index, and site.

\(b\) Transitions from Non-depressed df=2967; Transitions from Depressed df=1220
Treating Depression in Predominantly Low-Income Young Minority Women
A Randomized Controlled Trial

Jeanne Miranda, PhD
Joyce Y. Chung, MD
Bonnie L. Green, PhD
Janice Krupnick, PhD
Juned Siddique, MS
Dennis A. Revicki, PhD
Tom Belin, PhD

MAJOR DEPRESSION, A DISORDER WITH EARLY ONSET AND OFTEN CHRONIC COURSE, IMPOSES A HIGH INDIVIDUAL BURDEN OF PAIN, SUFFERING, AND DISABILITY.1 ETHNIC MINORITY PATIENTS ARE LESS LIKELY TO OBTAIN CARE FOR DEPRESSION THAN WHITE PATIENTS2 AND ARE LESS LIKELY TO RECEIVE APPROPRIATE TREATMENT WHEN THEY DO SEEK CARE.3,4 THESE DISPARITIES ARE PARTIALLY DUE TO MINORITIES BEING LESS LIKELY TO BE INSURED,2 WHICH RESULTS IN POOR ACCESS TO CARE. IN THIS STUDY, WE EVALUATED WHETHER INTERVENTIONS THAT PROMOTE GUIDELINE-COncORDANT TREATMENTS FOR DEPRESSION AMONG YOUNG, PREDOMINANTLY MINORITY WOMEN WOULD IMPROVE RATES OF APPROPRIATE CARE AND CLINICAL AND FUNCTIONAL OUTCOMES.

Effective treatments for major depression include antidepressant medications and psychotherapies.5,6 MOST US PSYCHIATRISTS FAVOR THE SELECTIVE SEROTONIN REUPTAKE INHIBITORS FOR FIRST-LINE TREATMENT.7 TWO BRIEF, STRUCTURED PSYCHOtherAPIES, COGNITIVE BEHAVIORAL THERAPY (CBT)8 AND INTERPERSONAL PSYCHOThERAPY,9 HAVE DEMONSTRATED EFFECTIVENESS IN TREATING PSYCHIATRIC PATIENTS.10-14 HOWEVER, EFFICACY TRIALS HAVE BEEN CONDUCTED ON Predominantly white patients in academic psychiatric settings. According to a review conducted as part of a surgeon general’s report,2 the studies forming the evi-

Context Impoverished minority women experience a higher burden from depression than do white women because they are less likely to receive appropriate care. Little is known about the effectiveness of guideline-based care for depression with impoverished minority women, most of whom do not seek care.

Objective To determine the impact of an intervention to deliver guideline-based care for depression compared with referral to community care with low-income and minority women.

Design, Setting, and Participants A randomized controlled trial conducted in the Washington, DC, suburban area from March 1997 through May 2002 of 267 women with current major depression, who attended county-run Women, Infants, and Children food subsidy programs and Title X family planning clinics.

Outcomes Hamilton Depression Rating Scale measured monthly from baseline through 6 months; instrumental role functioning (Social Adjustment Scale) and social functioning (Short Form 36-Item Health Survey) measured at baseline and 3 and 6 months.

Interventions Participants were randomly assigned to an antidepressant medication intervention (trial of paroxetine switched to bupropion, if lack of response) (n=88), a psychotherapy intervention (8 weeks of manual-guided cognitive behavior therapy) (n=90), or referral to community mental health services (n=89).

Results Both the medication intervention (P<.001) and the psychotherapy intervention (P=.006) reduced depressive symptoms more than the community referral did. The medication intervention also resulted in improved instrumental role (P=.006) and social (P=.001) functioning. The psychotherapy intervention resulted in improved social functioning (P=.02). Women randomly assigned to receive medications were twice as likely (odds ratio, 2.04; 95% confidence interval, 0.98-4.27; P=.057) to achieve a Hamilton Depression Rating Scale score of 7 or less by month 6 as were those referred to community care.

Conclusions Guideline-concordant care for major depression is effective for these ethnically diverse and impoverished patients. More women engaged in a sufficient duration of treatment with medications compared with psychotherapy, and outcome gains were more extensive and robust for medications.

JAMA. 2003;290:57-65

©2003 American Medical Association. All rights reserved.
dence base for the American Psychiatric Association guidelines for depression care included 3860 participants; only 27 were identified as African American and none were of any Latino descent. These trials also included primarily upper socioeconomic status populations. For example, the National Institute of Mental Health Treatment of Depression Collaborative Research Program11 sample included 40% college graduates. These science implementation studies have yielded a large gap in the knowledge base underlying treatment of low-income and minority women with depression, which could contribute to clinician uncertainty about the value of treating this group. This study addresses this gap in the literature by examining outcomes of care for low socioeconomic status and minority women.

Recent studies have examined care delivery systems for treating depression among primary care patients. Although these studies examined a more diverse population than smaller efficacy studies, the participants were predominantly insured and were current patients of a primary health care organization.15-22 An early finding in this literature suggests that management interventions in primary care may be particularly beneficial for minority individuals,23 but this study examined individuals who were already engaged in care. Young minority women, a group at high risk for depression,2 are often uninsured24 and many use county family planning and nutritional programs as a primary location of care. In addition, trauma and posttraumatic stress disorder (PTSD) are more prevalent in these subgroups of women with the lowest education and income.25 New literature suggests that care management strategies are needed to engage low-income minority individuals in treatment.26 Our intervention includes these enhancements. This study broadens the base of our current knowledge by testing whether guideline-based care for depression improves depressive symptoms and functioning in low-income minority women compared with women receiving community referrals for care.

METHODS

Women Entering Care was a randomized controlled trial of treatment for major depression in low-income women who received county health and welfare services in the Washington, DC, area suburban counties of Prince Georges and Montgomery, Md, and Arlington and Alexandria, Va. The study was reviewed and approved by the institutional review boards of George-town University, the University of California at Los Angeles, and the state of Maryland. All participants gave written informed consent for screening and separate consent for random assignment to care and for medical record abstraction.

Sample Participants

We screened 16 286 women: 11 151 in Prince Georges County and 3034 in Montgomery County, Md, and 2101 in Arlington and Alexandria, Va. Women were screened in Women, Infants, and Children food subsidy programs that target low-income pregnant and postpartum women and their children (≤5 years of age). Women were also screened in county-run Title X family planning clinics that are funded by national grants for comprehensive family planning services for young and low-income women. We screened 10043 women attending Women, Infants, and Children program services, 5017 attending family planning services, 1144 bringing children to pediatric services for low-income families, and 82 living in subsidized housing projects or attending programs for county welfare recipients.

Women were screened for major depressive disorder (MDD) while they were at the service setting. Screenings were conducted from March 1997 through December 2001, with the treatment ending in May 2002. Refusal rates for screening were very low. Figure 1 depicts the screening and enrollment process for the study. To determine whether there were cultural differences in response to interventions, we recruited 3 distinct cultural groups: black women born in the United States, Latinas born in Latin America, and white women born in the United States. Of the 16 286 women screened for MDD, 13 975 were eligible based on self-reported ethnicity and country of origin. Approximately 11% of these women screened positive for MDD based on the Primary Care Evaluation of Mental Disorders,27 and 36% met exclusion criteria, including bereave-

Engagement Into Care

Following the initial clinic screenings, we attempted to contact women who agreed to participate in the telephone diagnostic portion of the study. Participants were contacted by telephone a mean (SD) of 4.1 (4.4) times before completing the diagnostic interview by telephone. Despite repeated attempts, 345 participants (33.9%) were not reachable. Of those participants completing the screening, 427 (42%) completed a structured psychiatric diagnostic telephone interview (Composite International Diagnostic Interview28) and met criteria for the study. Of those participants not eligible, the majority did not meet criteria for major depression.

Clinicians contacted the women a mean (SD) of 7.8 (9.8) times to encourage them to attend an initial clinical interview; consent for random assignment was obtained at the beginning of the interview. Eleven women refused randomization. The clinicians were unaware of a participant’s random assignment until the end of the interview, at which time the computer-generated random treatment assignment was re-
vealed by the interviewer telephoning a newly recorded message. A total of 267 Composite International Diagnostic Interview–eligible women (63%) completed the clinical interview, consented to treatment, and were randomized into the intervention trial. Women were randomized as follows: antidepressant medications administered by a primary care nurse practitioner in consultation with a psychiatrist (n=88), CBT conducted by a psychologist (n=90), or referral to community mental health services (n=89). Of those randomized, 117 were black women, 134 were Latina women, and 16 were white women. Among those women randomized, rates of obtaining care varied by randomized assignment. Response to follow-up interviews was generally high, with 80% completing 3 or more of 6 follow-up interviews.

**Interventions**

**Education Meetings.** Participants assigned to receive either CBT or medications were given both time and opportunity for discussion of depression and treatment during education meetings with the clinician overseeing their care before beginning treatment. Up to 4 education meetings could be scheduled to help participants decide about initiating treatment.

**Medication.** Women assigned to the medication intervention were treated by primary care nurse practitioners supervised by a board-certified psychiatrist (J.Y.C.). The duration of the medication intervention was 6 months, in line with Agency for Healthcare Research and Quality guidelines for the acute and maintenance phases of depression treatment.29 Women were initially treated with paroxetine, which was prescribed according to a written dosing protocol. Adjustments in dosage were based on changes in Hamilton Depression Rating Scale (HDRS)30 scores and adequate time for medication effect. The range of paroxetine dose was 10 to 50 mg daily with a mean dose of 30 mg. Clinicians also assessed participants at face-to-face meetings by using the Clinical Global Impression Scale.31 If a patient did not tolerate paroxetine because of adverse effects or, if by the ninth week, the patient did not show a significant clinical response despite dose adjustments, buproprion, an antidepressant with a different presumed mechanism of action as well as a different adverse effect profile, was ad-

---

**Figure 1. Participant Flow in Women Entering Care Study**

<table>
<thead>
<tr>
<th>267 Randomized</th>
</tr>
</thead>
<tbody>
<tr>
<td>16286 Women Screened</td>
</tr>
<tr>
<td>13975 Culturally Eligible</td>
</tr>
<tr>
<td>1583 Had Positive PRIME-MD Result for MDD</td>
</tr>
<tr>
<td>1017 Screened Eligible</td>
</tr>
<tr>
<td>427 Met Criteria for MDD Using CIDI</td>
</tr>
<tr>
<td>88 Assigned to Receive Antidepressant Medication</td>
</tr>
<tr>
<td>90 Assigned to Receive CBT</td>
</tr>
<tr>
<td>89 Referred to Community Health Services</td>
</tr>
<tr>
<td>67 Received ≥6 wk of Guideline Concordant Medication Therapy</td>
</tr>
<tr>
<td>60 Completed ≥3 Assessment Visits</td>
</tr>
<tr>
<td>21 Received &lt;6 wk of Guideline Concordant Medication Therapy</td>
</tr>
<tr>
<td>15 Completed ≥3 Assessment Visits</td>
</tr>
<tr>
<td>32 Received ≥6 CBT Sessions</td>
</tr>
<tr>
<td>30 Completed ≥3 Assessment Visits</td>
</tr>
<tr>
<td>58 Received &lt;6 CBT Sessions</td>
</tr>
<tr>
<td>38 Completed ≥3 Assessment Visits</td>
</tr>
<tr>
<td>88 Included in Analysis</td>
</tr>
<tr>
<td>90 Included in Analysis</td>
</tr>
<tr>
<td>89 Included in Analysis</td>
</tr>
<tr>
<td>2311 Excluded</td>
</tr>
<tr>
<td>1345 Non-US-Born Black</td>
</tr>
<tr>
<td>345 Asian/American Indian</td>
</tr>
<tr>
<td>67 Non-US-Born White</td>
</tr>
<tr>
<td>184 US-Born Latina</td>
</tr>
<tr>
<td>178 Multiple Ethnicities</td>
</tr>
<tr>
<td>184 Unknown</td>
</tr>
<tr>
<td>566 Excluded</td>
</tr>
<tr>
<td>35 Bereaved</td>
</tr>
<tr>
<td>12 Had Alcohol or Drug Problems</td>
</tr>
<tr>
<td>185 Planned Pregnancy</td>
</tr>
<tr>
<td>225 Breastfeeding</td>
</tr>
<tr>
<td>109 In Mental Health Treatment</td>
</tr>
<tr>
<td>160 Excluded</td>
</tr>
<tr>
<td>149 No Clinical Interview</td>
</tr>
<tr>
<td>11 Refused Assignment</td>
</tr>
</tbody>
</table>

PRIME-MD indicates Primary Care Evaluation of Mental Disorders; CIDI, Composite International Diagnostic Interview; MDD, major depressive disorder; CBT, cognitive behavioral therapy. Community provider could be any mental health organization.

©2003 American Medical Association. All rights reserved.
ministered. Participants treated with medications were observed in person every 2 weeks until a stable dose was achieved, and then every 4 weeks thereafter. The nurses scheduled weekly telephone calls to assess adverse effects, adherence, and treatment effects.

Cognitive Behavioral Therapy. Women assigned to the psychotherapy intervention were treated by experienced psychotherapists supervised by a licensed clinical psychologist with CBT expertise. The psychotherapy intervention consisted of 8 weekly sessions, either in group or individually, depending on feasibility of women attending group. Sessions were conducted at the county clinics, Women Entering Care offices and, if necessary because of childcare or elder-care constraints, in a participant’s home. Participants were each given a CBT manual that they kept after treatment was concluded. The manual-guided treatment was adapted from 12-session patient and therapist manuals developed for low-income English- and Spanish-speaking medical patients.32–33 The manual was also shortened to 8 sessions by including more topics per session and modified to be more sensitive to the issues of young women and those with histories of interpersonal trauma. Therapists received training in understanding PTSD and trauma. The therapy protocol involved homework and daily monitoring, with a focus on cognitive management of mood, engaging in pleasant activities, and improving relationships with others. Improvement was assessed by using the HDRS and Beck Depression Inventory.34 If after 8 weeks of treatment a participant’s scores were still elevated, an additional 8-week treatment was offered.

Referral to Community Care. Women assigned to community referral as usual were educated about depression and mental health treatments available in the community. The clinician offered to make an appointment for the woman at the end of the clinical interview to facilitate the referral and to speak with the mental health care clinician. Approximately one quarter of the women declined referral. Referred participants were contacted to encourage them to attend the intake appointment for care.

Cultural Adaptation of Care. Bilingual providers treated all Spanish-speaking women. All written materials, including psychotherapy manuals, were available in Spanish. Of the 6 psychotherapists, 1 was black and 3 were Spanish speaking; none of the 4 nurse practitioners were black, and 2 were Spanish speaking. All psychotherapists and nurse practitioners had extensive experience with and commitment to treating low-income and minority patients.

Measures and Data Used
All measures were read to participants based on the high proportion of women who had not finished high school. Screening interviews assessed age, marital status, employment status, ethnicity, country of birth, and level of education. The mood disorders section of the Primary Care Evaluation of Mental Disorders52 was used to identify women at high risk for MDD. Alcohol problems were ascertained using the 5-item TWEAK (Tolerance, Worry about drinking, Eye-opener drinks, Amnesia, Cut down on drinking) designed for obstetric-gynecology clinic patients.33 Substance abuse problems were assessed by using items from the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (SCID).36

To determine whether women met criteria for MDD, the Composite International Diagnostic Interview28 was administered by telephone, assessing 12-month major depressive episode, alcohol abuse or dependence, drug abuse or dependence, and lifetime mania and psychosis. To be eligible for the study, participants needed to have current major depression and a negative score for mania, psychosis, or past month alcohol or drug abuse or dependence.

During the initial in-person interview with clinicians, women completed a measure of trauma history (Stressful Life Event Screening Questionnaire).37 Up to 3 events were evaluated for current PTSD by using the PTSD module of the SCID.36 Women were reimbursed $15 for completing each interview.

During baseline and follow-up telephone interviews, participants completed a structured version of the HDRS each month for 6 months. Functional outcomes, including instrumental role functioning (employee or homemaker role assessed by the Social Adjustment Scale interview) and social functioning (measured by the Short Form 36-Item Health Survey) were completed at baseline and months 3 and 6. Telephone interviewers were blinded to random assignment of the participants, and women were reimbursed $10 for completing each interview.

Data Analyses
Clinical outcomes were determined through a mixed-effects repeated-measures analysis comparing mean depression symptom and functioning scores across assigned treatment groups over successive time periods. For depressive symptom outcomes that were measured monthly, we fit models using outcomes from month 1 through 6 while controlling for the baseline values. For instrumental role and social functioning outcomes, we fit models using the baseline, month 3, and month 6 scores. In every model, we allowed for the within-participant correlations over time by using a random intercept or random intercept and slope covariance structure. We examined the effect of treatment over time, as well as the interaction of treatment and ethnicity.

We also examined the interaction of treatment and PTSD on outcomes. In this study, the assigned clinician completed the PTSD measure during an initial clinical interview. All other measures were conducted by telephone by trained assessors. Despite comprehensive training on the standardized instrument (SCID), we found that nurse practitioners were significantly more likely to diagnose PTSD (61% of women interviewed) compared with psychologists (34%, P < .001), suggesting that nurses used less stringent criteria in
judging presence of symptoms. Because of the importance of PTSD in this population, we performed analyses of medication vs community referral among participants diagnosed with PTSD by nurses and separate analyses for psychotherapy participants vs community referral among participants diagnosed by psychologists. There was no interaction of PTSD and treatment for either group, and PTSD was not included in other analyses.

To examine the robustness of our findings, we performed intent-to-treat unadjusted analysis of variance and permutation comparisons of 6-month outcomes. Because these 2 analytic approaches produce similar findings, we report only analysis of variance results. In addition, we replicated the major longitudinal analyses by using multiple imputed data sets. With the exception of income (n=10), no baseline data were missing. The extent of missing data at follow-up was similar across the 3 randomized groups. In total, 11 women (4%) missed all but the baseline interview, and 64 (24%) missed at least the last 2 interviews. Of all possible interviews, 29% were missing. To impute item-level missing data, we used an extended hot-deck multiple imputation technique that modifies the predictive mean-matching method. We used an approximate Bayesian bootstrap method to impute unit-level missing data. Imputations were produced separately for each intervention and the community referral groups. SAS statistical software version 8.2 was used for all analyses (SAS Institute, Cary, NC), and P<.05 was considered significant.

**RESULTS**

**Participants**

The participants in this study were primarily working poor Latina and black women. The mean (SD) age of participants was 29.3 (7.9) years (Table 1). Less than half were married or living with a partner, and 34.1% had never married. More than one third of the women (37.1%) had not graduated from high school or received a general equivalency diploma, and only 6.7% had completed college. The average woman in this study had 2 or more children, and 64.8% of women were medically uninsured.

The sample was predominantly poor. More than half (60%) of the women lived at or below the federal guidelines for poverty (during the years from 1997 to 2001). Furthermore, another 34.2% were near poor (between 100%-200% of poverty guidelines). In addition, the women in this study experienced extremely high rates of trauma exposure. More than one third had been raped and/or experienced child abuse, and 49.2% had experienced domestic violence. High rates of PTSD were also found.

### Table 1. Characteristics of Intervention and Community Referral Participants*

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N = 267)</th>
<th>Medication (n = 88)</th>
<th>Cognitive Behavioral Therapy (n = 90)</th>
<th>Referral to Community Mental Health Services (n = 89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD), y</td>
<td>29.3 (7.9)</td>
<td>28.7 (6.6)</td>
<td>29.8 (7.9)</td>
<td>29.5 (9.1)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or living with partner</td>
<td>124 (46.4)</td>
<td>43 (48.9)</td>
<td>40 (44.4)</td>
<td>41 (46.1)</td>
</tr>
<tr>
<td>Widowed or separated/divorced</td>
<td>52 (19.5)</td>
<td>17 (19.3)</td>
<td>22 (24.4)</td>
<td>13 (14.6)</td>
</tr>
<tr>
<td>Never married</td>
<td>91 (34.1)</td>
<td>28 (31.8)</td>
<td>28 (31.1)</td>
<td>35 (39.3)</td>
</tr>
<tr>
<td>No. of children, mean (SD)</td>
<td>2.3 (1.4)</td>
<td>2.2 (1.2)</td>
<td>2.2 (1.5)</td>
<td>2.4 (1.6)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>99 (37.1)</td>
<td>37 (42.0)</td>
<td>27 (30.0)</td>
<td>35 (39.3)</td>
</tr>
<tr>
<td>High school or GED</td>
<td>87 (32.6)</td>
<td>31 (35.2)</td>
<td>29 (32.2)</td>
<td>27 (30.3)</td>
</tr>
<tr>
<td>Some trade or college</td>
<td>63 (23.6)</td>
<td>15 (17.1)</td>
<td>28 (29.8)</td>
<td>22 (24.7)</td>
</tr>
<tr>
<td>College graduate</td>
<td>18 (6.7)</td>
<td>5 (5.7)</td>
<td>8 (8.9)</td>
<td>5 (5.6)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>117 (43.8)</td>
<td>34 (38.6)</td>
<td>41 (45.6)</td>
<td>42 (47.2)</td>
</tr>
<tr>
<td>White</td>
<td>16 (6.0)</td>
<td>6 (6.8)</td>
<td>6 (6.7)</td>
<td>4 (4.5)</td>
</tr>
<tr>
<td>Latina</td>
<td>134 (50.2)</td>
<td>48 (54.6)</td>
<td>43 (47.8)</td>
<td>43 (48.3)</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninsured</td>
<td>173 (64.8)</td>
<td>55 (62.5)</td>
<td>58 (64.4)</td>
<td>60 (67.4)</td>
</tr>
<tr>
<td>Medical assistance</td>
<td>40 (15.0)</td>
<td>14 (15.9)</td>
<td>12 (13.3)</td>
<td>14 (15.7)</td>
</tr>
<tr>
<td>Private</td>
<td>54 (20.2)</td>
<td>19 (21.6)</td>
<td>20 (22.2)</td>
<td>15 (16.9)</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working or looking for work</td>
<td>219 (82.0)</td>
<td>69 (78.4)</td>
<td>76 (84.4)</td>
<td>74 (83.2)</td>
</tr>
<tr>
<td>Not working or disabled</td>
<td>48 (18.0)</td>
<td>19 (21.6)</td>
<td>14 (15.6)</td>
<td>15 (16.9)</td>
</tr>
<tr>
<td>Poverty†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below federal poverty</td>
<td>149 (60.0)</td>
<td>48 (57.1)</td>
<td>48 (56.5)</td>
<td>53 (60.2)</td>
</tr>
<tr>
<td>Near poor (100%-200% poverty guidelines)</td>
<td>88 (34.2)</td>
<td>33 (39.3)</td>
<td>27 (31.8)</td>
<td>28 (31.8)</td>
</tr>
<tr>
<td>Not impoverished</td>
<td>20 (7.8)</td>
<td>3 (3.6)</td>
<td>10 (11.8)</td>
<td>7 (8.0)</td>
</tr>
<tr>
<td>Experience of trauma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life-threatening illness</td>
<td>51 (19.5)</td>
<td>23 (26.4)</td>
<td>12 (13.6)</td>
<td>16 (18.4)</td>
</tr>
<tr>
<td>Life-threatening accident</td>
<td>36 (13.6)</td>
<td>11 (12.5)</td>
<td>13 (14.6)</td>
<td>12 (13.6)</td>
</tr>
<tr>
<td>Robbed or mugged</td>
<td>48 (19.1)</td>
<td>16 (18.6)</td>
<td>16 (18.3)</td>
<td>18 (20.5)</td>
</tr>
<tr>
<td>Family/friend died violently</td>
<td>88 (33.2)</td>
<td>26 (29.5)</td>
<td>38 (42.7)</td>
<td>24 (27.3)</td>
</tr>
<tr>
<td>Raped</td>
<td>99 (37.2)</td>
<td>33 (37.5)</td>
<td>32 (35.6)</td>
<td>34 (38.6)</td>
</tr>
<tr>
<td>Child abuse</td>
<td>89 (33.5)</td>
<td>32 (36.4)</td>
<td>29 (32.2)</td>
<td>28 (31.8)</td>
</tr>
<tr>
<td>Domestic violence</td>
<td>129 (49.2)</td>
<td>45 (51.7)</td>
<td>49 (54.0)</td>
<td>37 (42.1)</td>
</tr>
<tr>
<td>Threatened with a weapon</td>
<td>56 (21.3)</td>
<td>16 (18.2)</td>
<td>20 (22.7)</td>
<td>20 (23.0)</td>
</tr>
<tr>
<td>Posttraumatic stress disorder§</td>
<td>127 (47.7)</td>
<td>53 (60.2)</td>
<td>31 (34.4)</td>
<td>43 (48.9)</td>
</tr>
<tr>
<td>Hamilton Depression Rating Scale, mean (SD)</td>
<td>16.9 (5.2)</td>
<td>18.0 (5.1)</td>
<td>16.3 (5.2)</td>
<td>16.5 (5.2)</td>
</tr>
</tbody>
</table>

Abbreviation: GED, general equivalency diploma.

*Data are No. (%) unless otherwise specified.

†Ten participants did not provide income information.

§One participant did not receive the posttraumatic stress disorder measure.

§§Differences in posttraumatic stress disorder among treatment groups appear to result from nurse vs psychologist diagnoses (P = .002).
Recruitment and Participation in Care

Given the multiple demands on these working-poor women, outreach was an essential part of this study. Following the clinical interview, nurse practitioners spoke with women a mean (SD) of 8.8 (11.2) times to enable a first medication visit. The psychologists spoke with patients a mean (SD) of 10.2 (12.2) times before they attended a psychotherapy visit. Women who were reluctant to begin medications or psychotherapy could attend education sessions to teach them about depression treatment before agreeing to enter care. Eighty-five women (96%) attended a mean (SD) of 1.89 (0.91) educational sessions before beginning medications, and 60 women (67%) attended 2.37 (1.76) educational sessions before beginning psychotherapy. Of the women who were randomly assigned to medication treatment, 18 (20%) were switched to buproprion. Of the women assigned to psychotherapy, 15 (17%) received an additional 8 sessions of CBT. Transportation to care visits and child care funds were provided to enable women to participate in the 2 interventions. Child care was provided by either reimbursing women $10 per hour of treatment so that they could pay someone they knew for child care or when they were unable to identify a potential child care provider, child care was provided at the treatment site by study research assistants. Transportation was provided through taxi vouchers, a prepaid hired van, or reimbursement of subway or bus costs, depending on the availability of transportation at a particular site.

Participation in care was low for women in the community referral condition. After referral to a specific community provider, 74 (83%) of the women offered community referral failed to attend even 1 session. Only 8 women (9%) attended 4 or more sessions of community care. Of those randomly assigned to receive medications, 67 (75%) received 9 or more weeks of a guideline-concordant course of medications, and 40 (45%) received a guideline-concordant course for 24 or more weeks. Women were less likely to complete an adequate course of CBT. Of those randomly assigned to receive psychotherapy, 48 (53%) received 4 or more CBT sessions and 32 (36%) received 6 or more sessions. Likelihood of receiving an adequate course of treatment did not differ by ethnicity.

Longitudinal Outcomes

Results of repeated-measures analyses assessing change across time revealed a significant decrease in depressive symptoms for women in the medication (\(P < .001\)) and psychotherapy (\(P = .006\)) interventions compared with the community referral group when measured by the HDRS (FIGURE 2). There were no significant interactions in depression outcomes between the interventions and ethnicity. The women randomly assigned to receive medication and psychotherapy reported improved functioning compared with those referred to community care. Results on instrumental role functioning revealed a significant impact of the medication intervention (\(P = .006\)), but not the psychotherapy intervention (\(P = .38\)) when compared with the group referred to community care. Both medication (\(P = .001\)) and psychotherapy (\(P = .02\)) interventions were effective in improving social functioning vs community referral. The women randomly assigned to receive medications were twice as likely (odds ratio, 2.04; 95% confidence interval, 0.98-4.27; \(P = .057\)) to achieve an HDRS score of 7 or less by month 6 as were those referred to community care. By month 6, 44.4% of medication, 32.2% of psychotherapy, and 28.1% of community referral patients had achieved an HDRS score of 7 or less.

Comparisons were made between adjusted mean scores of those in the intervention groups and those in the community referral group at months 1, 3, and 6 (TABLE 2). Women in the medication intervention had lower depressive symptoms at months 3 and 6 and higher instrumental role and social functioning by month 6 than did those referred to community care. Women assigned to receive psychotherapy had lower depressive symptoms by month 6 and improved social functioning at month 6 compared with the referred participants. Women in the psychotherapy group did not differ from the referred controls on the instrumental role functioning measure. Women receiving medication reported lower depressive symptoms and better instrumental role functioning than did those receiving psychotherapy at month 6.

Examination of Robustness of Results

To examine the robustness of our longitudinal findings, we reexamined the major analyses using intent-to-treat analysis of variance test at 6 months, with no covariates. We replicated our finding that the medication intervention was superior to community referral in decreasing depressive symptoms (\(P = .04\)), improving instrumental role functioning (\(P = .01\), and social functioning (\(P = .01\). The psychotherapy intervention was not superior to community referral in decreasing depressive symptoms (\(P = .32\) or improving role functioning (\(P = .58\), but did result in improved social functioning (\(P = .06\).

As an additional test of robustness, we used 5 imputed data sets to ensure that
our results were not due to missing data. The results for the medication intervention compared with community referral were replicated for decreasing depressive symptoms (P = .04) and improving instrumental role (P = .009) and social (P = .003) functioning. With the exception of improving social functioning (P = .03), there were no significant differences between the psychotherapy participants and the community referral group for the imputed data.

### Analysis of Treatment Received

To examine the likely outcomes of current practice guideline treatment in this population, we compared those women in the medication group who received at least 9 weeks of guideline-concordant treatment (n = 67) and those in the psychotherapy group who received at least 6 CBT visits (n = 32) with those in the community referral group who did not receive any mental health treatment (n = 74). Guideline medications (P < .001) and psychotherapy (P = .003) were significantly more effective than no treatment in decreasing depressive symptoms. There were no interactions between treatment and ethnicity. Medication (P = .003) was superior in improving instrumental role functioning compared with no treatment; psychotherapy was not (P = .07). The medication (P < .001) and the psychotherapy (P = .003) interventions resulted in improved social functioning compared with no treatment.

Comparisons were made between mean scores at months 1, 3, and 6 for women receiving guideline treatment compared with women receiving community referral with no treatment (Table 3). To determine the impact of guideline care above care received in the community, we also compared women receiving guideline medication or psychotherapy treatment with community referral individuals who entered community care (n = 15). Guideline medications were more effective than community care at month 6 for reducing depressive symptoms (P < .001), increasing instrumental role functioning (P = .01) and social functioning (P < .001). Guideline psychotherapy was more effective than community care at 6 months for reducing depressive symptoms (P = .01) and social functioning (P = .006), but did not improve instrumental role functioning (P = .78).

### Comment

These results demonstrate the effectiveness of the study interventions for depression when provided to low-income and minority women. The interventions included intensive outreach, child care and transportation to care when needed, and encouragement to comply with evidence-based medication or psychotherapy treatment. As a result of this care, women not only achieved lower levels of depressive symptoms, but they also achieved higher levels of instrumental role and social functioning.

The results for community referral show that few impoverished women engage in community care that is available to them. Although we could obtain appointments for care, very few women actually attended such care. Without the outreach, child care and transportation, and flexible scheduling of care offered, few impoverished women are likely to receive appropriate treatment for depression.

Engagement of women in care was demanding throughout this study. We were unable to obtain telephone diagnostic interviews for nearly one third of the women who screened positive for depression, despite persistent attempts. Of those who were interviewed, 26% who had screened positive did not meet criteria for major depression. Many reported that their mood was low at
screening because of life circumstances but their mood had subsequently improved. Once diagnosed, our clinicians spent considerable time gaining the trust of the women, often through telephone contact, before they were comfortable receiving care.

We found no ethnic differences in response to care. Although our sample of white women was extremely small, these results argue against concerns that evidence-based treatment may not be effective for black and Latina women. Evidence-based interventions appear effective for poor and minority women if they are given support to overcome barriers to care. The interventions were moderately tailored for each ethnic population, providing care in Spanish when appropriate, and using professionals sensitive to low-income and minority populations. With these sensitivities, evidence-based treatment for depression appears effective for diverse populations.

Results of this study suggest that medication interventions may be more effective for low-income and minority women than are psychotherapy interventions. More women engaged in a sufficient duration of guideline treatment with medications compared with psychotherapy, and outcomes of care were more extensive and robust for medications. The nurse practitioners may be more accustomed to providing patient outreach than are psychotherapists. This finding could also reflect the lower time demands on patients of taking medications vs attending psychotherapy. Also, the results of psychotherapy may be slower to unfold.

Several limitations of this study should be considered. First, all measures were self-report, although the study measures have been used extensively in previous depression studies. Second, our measure of PTSD appears to be influenced by type of clinician completing the diagnosis. Nonetheless, presence of PTSD did not appear to lessen the impact of depression care. Third, nearly half of the women who screened positive for depression did not follow-up with diagnostic interviews, and others claimed to no longer be depressed when interviewed later. Thus, the extent to which this sample represents low-income young women in general is not known. Fourth, all women in the study received monthly telephone interviews to assess study outcomes. According to their self-reports, the women experienced these calls as helpful. The impact of multiple contacts for obtaining study measures could attenuate the results of the interventions compared with community referral. Finally, the impact of evidence-based treatment without the enhancements necessary to bring women to treatment cannot be determined in this study.

We conclude that our interventions, including medications and psychotherapy consistent with depression-treatment guidelines, are superior to community referral for low socioeconomic status young women. The health policy implications of this study are important. Engaging impoverished minority women in care is demanding, but outcomes are clearly beneficial. Engaging women through trusted providers could prove easier. Our results demonstrate that treating depression in this population has clear advantages, both in terms of reducing personal suffering and improving the ability of these young women to function.

Author Contributions: Study concept and design: Miranda, Chung, Green, Krupnick, Revicki. Acquisition of data: Miranda, Chung, Krupnick. Analysis and interpretation of data: Miranda, Chung, Green, Siddique, Revicki, Bellin. Drafting of the manuscript: Miranda, Chung, Green, Siddique, Bellin.

Critical revision of the manuscript for important intellectual content: Miranda, Chung, Green, Krupnick, Revicki.
REFERENCES


One-Year Outcomes of a Randomized Clinical Trial Treating Depression in Low-Income Minority Women

Jeanne Miranda
University of California, Los Angeles

Juned Siddique and Tom Belin
University of California, Los Angeles

Bonnie L. Green, Janice L. Krupnick, and Joyce Chung
Georgetown University Medical Center

Dennis Revicki
Georgetown University Medical Center and MEDTAP International

This study examines 1-year depressive symptom and functional outcomes of 267 predominantly low-income, young minority women randomly assigned to antidepressant medication, group or individual cognitive–behavioral therapy (CBT), or community referral. Seventy-six percent assigned to medications received 9 or more weeks of guideline-concordant doses of medications; 36% assigned to psychotherapy received 6 or more CBT sessions. Intent-to-treat, repeated measures analyses revealed that medication (p < .001) and CBT (p = .02) were superior to community referral in lowering depressive symptoms across 1-year follow-up. At Month 12, 50.9% assigned to antidepressants, 56.9% assigned to CBT, and 37.1% assigned to community referral were no longer clinically depressed. These findings suggest that both antidepressant medications and CBT result in clinically significant decreases in depression for low-income minority women.

Keywords: depression, clinical trials, ethnic-minority women, cognitive–behavioral therapy

Ethnic minority and poor individuals are less likely to receive treatment, particularly guideline-informed care, for major depressive disorder than are White and middle-class individuals (U.S.

Jeanne Miranda and Juned Siddique, Health Services Research Center, Neuropsychiatric Institute, University of California, Los Angeles; Bonnie L. Green, Janice L. Krupnick, and Joyce Chung, Department of Psychiatry, Georgetown University Medical Center; Tom Belin, Health Services Research Center, Neuropsychiatric Institute, University of California, Los Angeles, and Department of Biostatistics, University of California at Los Angeles School of Public Health; Dennis Revicki, Department of Psychiatry, Georgetown University Medical Center, and Center for Health Outcomes Research, MEDTAP International, Bethesda, MD.

This work was funded by National Institute of Mental Health (NIMH) Grant R01 MH56864; by the John D. and Catherine T. MacArthur Foundation; and by NIMH Grant 5F30MH068639, National Center on Minority Health and Health Disparities Grant 5P20MD000148, and National Institute on Aging Grant 5:30A6021684. Paroxetine was provided by GlaxoSmithKline.

We thank Pat Heiber for her work as project coordinator; Amy Cha, Sharon Gearhart, Lisa Hooper, and Cliff Willis for grant and data management; and Anne Riley, Mary Jo Coiro, Marina Broitman, and Lori Frank for critical input and collaboration. The project would not have been possible without the dedication and skill of our clinicians (Jan Gagen, Lois Wessel, Jacqueline Cadena, Laura Kohn, Ana Daroowalla, Elisha Friedman, Karin Focke, Teresa Tudder, Elissa Emerson) and key staff/interviewers (Adrianna Ortiz and Sarah Conron).

Correspondence concerning this article should be addressed to Jeanne Miranda, Department of Psychiatry and Biobehavioral Sciences, Health Services Research Center, University of California, 10920 Wilshire Blvd., Suite 300, Los Angeles, CA 90024-6505. E-mail: JMMiranda@mednet.ucla.edu

Department of Health and Human Services [DHHS], 2001). This may be related to the fact that most depression treatment studies include primarily White and middle-class populations (DHHS, 2001), so that little is known about the usefulness of established treatments for more disadvantaged populations. Establishing the effectiveness of depression care among poor young women is particularly important because rates of depression are elevated in women, younger age cohorts, and those living in or near poverty (Kessler, Berglund, & Demler, 2003). Previously, we found that both an antidepressant medication intervention and a cognitive–behavioral psychotherapy intervention were significantly more effective than referral to community mental health care for lowering depressive symptoms and improving functioning among predominantly poor young minority women 6 months after their depression was identified (Women Entering Care [WE Care]; Miranda et al., 2003). In this article, we examine 1-year outcomes for the same sample of women.

Long-term outcomes of depression care are important to examine because, even among more advantaged populations, many individuals do not sustain remission of symptoms. For example, 40% of the recovered sample relapsed by 18-month follow-up in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Shea et al., 1992). The accumulated stressors and lack of resources for coping with these stressors may diminish the benefits of short-term depression care for poor young women even beyond that seen with more advantaged populations. In a study of public mental health sector outpatients receiving guideline care, only 26% were considered in remission at 1 year (Rush et al., 2004).
In the WE Care study, we examined the effectiveness of well-established, evidence-based interventions for our impoverished population. Most U.S. psychiatrists favor the selective serotonin reuptake inhibitors for first-line medication treatment (Olfson & Klerman, 1993), with treatment extended to at least 6 months to maintain clinical effectiveness (Agency for Healthcare Research and Quality, 1993). Cognitive–behavioral therapy (CBT) is also an effective treatment for major depression. A meta-analysis of six randomized trials of antidepressant medications, CBT, or interpersonal psychotherapy (Casacalenda, Perry, & Looper, 2002) found that remission rates in intent-to-treat analyses average 46% for medications, 46% for psychotherapy, and 24% for controls. DeRubeis et al., in their 2005 trial comparing paroxetine with cognitive therapy, found similar remission rates. In our WE Care study at 6 months, depression treatment outcomes showed that 44.4% of medication, 32.2% of psychotherapy, and 28.1% of community referral patients had remitted (Miranda et al., 2003).

Previous work has shown that CBT produces sustained clinical gains compared with antidepressant medications that are withdrawn after clinical response (Blackburn, Eunson, & Bishop, 1986; Evans et al., 1992; Kovacs, Rush, Beck, & Hollon, 1981; Miller, Norman, & Keitner, 1989; Shea et al., 1992; Simons, Murphy, Levine, & Wetzel, 1986). In a study of treatment responders, patients treated with cognitive therapy were more likely to have a sustained response during 12-month follow-up than those withdrawn from medications; and they were just as likely to have sustained response as patients who kept taking medications through the follow-up (Hollon et al., 2005), suggesting that CBT may have important advantages over the long term.

In the present study, we explored long-term outcomes of two active treatments compared with community referral for care for disadvantaged, young minority women. Although our earlier work suggested that medication had a slight advantage in the short term (6 months; Miranda et al., 2003), previous literature suggests that CBT might be of more benefit over a longer time period. Treatment for depression in advanced populations results in significant improvements in functional and work outcomes (Coulehan, Schulberg, Block, Madonia, & Rodriguez, 1997; Kocsis et al., 2002; Miller et al., 1998; Simon et al., 1996, 1999; Wells et al., 2000), with some finding long-term maintenance in these areas (Kocsis et al., 2002). We examined the impact of depression care on functioning of disadvantaged young women at 1-year follow-up as well. Specifically, we hypothesized that the psychotherapy intervention would result in fewer depressive symptoms and higher rates of remission at 12-month follow-up than the antidepressant intervention. Because functional outcomes appear closely tied to depressive symptoms, we predicted that the CBT intervention would also result in higher levels of functioning compared with the medication intervention.

Method

WE Care was a randomized, controlled trial of treatment for major depression in low-income women who received county health and welfare services. Women who agreed to random assignment were offered antidepressant medications by standard protocol, manual-driven CBT, or referral to community mental health care. The study was reviewed and approved by the institutional review boards of Georgetown University, University of California, Los Angeles, and the state of Maryland. All participants gave written informed consent for screening and separate consent for random assignment to care.

Sample

Details about participant selection, exclusion criteria, and treatments have been reported elsewhere (Miranda et al., 2003) and are summarized in Figure 1. We screened 16,286 women who received county health and welfare services in the Washington, DC, area suburban counties of Prince Georges and Montgomery, MD, and Arlington and Alexandria, VA. Of those screened, 13,975 were eligible based on ethnicity (African American, Latina, or White) and country of origin (United States for African Americans, Latin America for Latinas, and United States for Whites). Other groups were excluded so that ethnicity could be investigated empirically for its impact on outcome. Approximately 11% (1,585) of ethnically eligible women screened positive for major depressive disorder (MDD) on the Primary Care Evaluation of Mental Disorders (Spitzer et al., 1994), and 36% of those (566) were already receiving care or were excluded as inappropriate for our interventions (i.e., recently bereaved, current substance use disorders, plans to become pregnant or breast-feeding). Of those who screened eligible, 590 were subsequently excluded for either not meeting criteria for major depression (i.e., on the Composite International Diagnostic Interview [CIDI]; World Health Organization [WHO], 1997) and at least 1 week later reporting a score of 14 or greater on the Hamilton Depression Rating Scale (HDRS; Williams, 1988) or having comorbid disorders (schizophrenia, bipolar) that were likely to compromise response to our interventions. An additional 149 were excluded because they did not complete a final clinical interview. Only 11 women refused random assignment. Ultimately, 267 were eligible and randomly assigned to (a) antidepressant medications administered by a primary care nurse practitioner in consultation with a psychiatrist (n = 88), (b) CBT conducted by a psychologist (n = 90), or (c) referral to community mental health services (n = 89). Of these, 117 were African American, 134 were Latina, and 16 were White.

Demographic and clinical characteristics of the sample are presented in Table 1. There were no significant differences at baseline among the randomly assigned intervention groups. The sample was made up of young minority women, many of whom were uninsured and living below or near the poverty level. Women randomly assigned to medications reported somewhat higher levels of depressive symptoms at baseline than did the other two groups, a difference that neared significance (p = .06). On average, women reported an age of onset of depression of 23 to 24 years. About half the women were experiencing a mild to moderate episode and 47% a severe episode. Fewer than 20% of the women had experienced a previous episode of depression. Those with recurrent depression had experienced an average of about three episodes.

Outreach was an integral part of our intervention. Given the multiple demands on these working-poor women, outreach and help attending care was necessary to enable many to participate, including transportation and child-care funds. Women who screened positive for MDD were contacted by telephone a mean of 4.1 (± 4.4 SD) times before completing the CIDI (WHO, 1997) by telephone. Clinicians then contacted the women an average of 7.8 (± 9.8 SD) times to encourage them to attend an initial clinical interview.

Interventions

Educational meetings. The women in our trial were not seeking care for depression, and many seemed to be unaware that their feelings of sadness and distress were indicative of a clinical problem that could be treated. Therefore, we designed a flexible educational intervention for women who were not comfortable beginning treatment at the time of random assignment to the two active treatments. Women assigned to medications or psychotherapy were given up to four educational meetings.
with their provider to learn about depression and its treatment. These sessions also served as an opportunity for the women to build trust with their provider before accepting treatment. In the medication condition 85 women (96%) attended a mean of 1.89 (± 0.91 SD) sessions before beginning treatment. In the CBT condition, 60 women (67%) attended 2.37 (± 1.76 SD) educations sessions before beginning psychotherapy. This was not significantly different.

**Medication.** We adapted a standard medication protocol for our low-income population. Because outreach was necessary to engage women in care, nurse practitioners supervised by a board-certified psychiatrist (Joyce Chung) served as our clinicians. These nurses were accustomed to providing outreach to disadvantaged populations and were adept at engaging our target population in care. The medication protocol offered 6 months of antidepressant treatment. Women were initially treated with paroxetine, prescribed according to a written dosing protocol informed by clinical guidelines. Adjustments in dosage were based on changes in HDRS scores and adequate time for medication effects. The range of paroxetine doses were 10 to 50 mg daily (M = 30 mg). If the participant did not tolerate paroxetine or did not show a significant clinical response by 9 weeks despite dose adjustments, bupropion, an antidepressant with a different presumed mechanism of action as well as a different adverse effect profile, was administered.

---

**Figure 1.** Participant flow in WE Care study. CIDI = Composite International Diagnostic Interview; MDD = major depressive disorder; CBT = cognitive–behavioral therapy; PRIME-MD = Primary Care Evaluation of Mental Disorders. From “Treating depression in predominantly low-income young minority women: A randomized controlled trial,” by J. Miranda, J. Y. Chung, B. L. Green, J. Krupnick, J. Siddique, D. A. Revicki, and T. Belin, *JAMA, 290*, Figure 1. Copyright 2003, American Medical Association. Adapted with permission.
CBT. The psychotherapy treatment was administered by experienced psychotherapists previously trained in CBT and supervised by a licensed clinical psychologist with CBT expertise who conducted weekly group supervision of therapists to ensure adherence to the treatment. Therapy sessions were held at the county health clinics, WE Care offices, or, if necessary because of child- or elder-care responsibilities, in the participants’ homes. The manual guided-treatment was an 8-session modification of a 12-session intervention developed for low-income English- and Spanish-speaking medical patients (Munoz, Aguilar-Gaxiola, & Guzman, 1986; Munoz & Miranda, 1986). The manual was modified to include more topics per session and to be sensitive to the issues of young women, many of whom had histories of interpersonal trauma. Although we did not specifically elicit discussion of trauma during treatment, we consistently acknowledged these experiences and pointed out that they were likely related to current thoughts and behaviors. We also presented the CBT intervention as ways to cope with trauma-related problems. Women were offered group or individual care, depending on their schedules and possibilities for attending a group. Women who remained depressed (HDRS score > 7) after a course of CBT were offered 8 additional sessions.

Referral to community care. Women assigned to community referral were educated about depression and mental health treatments available in the community. The clinician offered to make an appointment for the woman at the end of the clinical interview to facilitate the referral and to speak with the mental health clinician. Approximately 25% of the women declined referral. Referred participants were contacted by the referring clinician within 1 or 2 weeks of referral to encourage them to attend the community care.

Measures

All outcome measures were orally administered, because a high proportion of women had not finished high school and may have been unable to read the materials. Screening interviews assessed age, marital status, employment status, ethnicity, country of birth, and level of education. To determine whether women met diagnostic criteria for MDD, the CIDI (WHO, 1997) was administered by telephone, assessing 12-month MDD, alcohol abuse or dependence, drug abuse or dependence, and lifetime mania and psychosis. To be eligible for the study, participants needed to have current MDD and to be negative for mania, psychosis, or alcohol or drug abuse or dependence in the past month.

Participants completed a structured version of the HDRS (Williams, 1988) by telephone at baseline, monthly for 6 months, and at Months 8, 10, and 12. Interviewers were trained to reliably administer this instrument by listening to and scoring tapes of clinician-administered HDRSs. In addi-
Data Analyses

The impact of random assignment to care on clinical and functional outcomes over 12 months was evaluated through a mixed-effects random intercept and slope repeated measures analysis comparing mean depression symptom and functioning scores across assigned treatment groups over successive time periods, with ethnicity, baseline HDRS scores, and baseline HDRS interacted with time as covariates. Mixed-effects models offer many advantages for analyzing longitudinal data in psychiatric research. The mixed-effects approach models changes over time at both the group and individual levels. A separate trend is estimated for each participant, based on available data, with missing data augmented by the individual’s complete data and information from all other participants in the sample. This allows for the fact that, over time, different individuals may respond to treatment differently. Finally, this model accounts for the fact that repeated measures within subjects are correlated, providing accurate estimates of the standard errors (Gibbons et al., 1993).

Mixed-effects models also use all available data. An end-point analysis, in which only the baseline and final measurement are considered, ignores a large amount of data that can provide a much better picture of how individuals change over time rather than snapshot images of the first and last time points. In addition, because time of last measurement can vary for each individual, time is effectively ignored in end-point analyses.

We fit models using data from baseline through Month 12. Because times varied for actual dates of interviews, our models used number of days since baseline as the time covariate. In addition, because depression scores differed by a nearly significant level (\( p = .06 \)) across treatment groups at baseline, all our models controlled for baseline depression and the Baseline Depression \( \times \) Time interactions. Depressive symptom outcomes were measured at Months 1, 2, 3, 4, 5, 6, 8, 10, and 12. Functional outcomes were measured at Months 1, 3, 6, and 12. We tested for differences among the three randomly assigned groups in the intercepts, linear slopes, and quadratic slopes. The linear slope measures continuous change in scores over time. We included the quadratic slope to determine whether the interventions simply cause steeper declines in depression scores (linear slope) or initial steeper declines in depression scores followed by a leveling out over time (quadratic slope).

To examine whether our interventions were more or less effective for our distinct cultural groups (African American, Latina, White), we examined the interaction of treatment and ethnicity. We had poor power for detecting minority versus White differences but could examine African American versus Latina differences. Because no Treatment \( \times \) Ethnicity interactions were significant, all results are presented for ethnic groups combined. We did, however, control for ethnicity in all of our models.

Participants assigned to psychotherapy were equally likely to attend group or individual care. We examined the impact of attending group versus individual psychotherapy and found no differences in outcome, so all results are presented for CBT care combined.

To ensure that our results were not dependent on a pattern of nonrandom missing data, we replicated the major longitudinal analyses by using last observation carried forward and by multiply imputed data sets. We present the multiply imputed results only, because both techniques resulted in similar findings, and imputed analyses provide a test of robustness of the findings regarding the influence of missing data. Further, last observation carried forward techniques assume that outcomes remain constant at the last observed value (unlikely to be true even for untreated depression), and this methodology does not adjust for uncertainty because of missing data that are accounted for in multiply imputed analyses. Baseline data were complete with the exception of 10 individuals missing income, 5 missing instrumental role functioning, and 2 missing social functioning. The extent of missing data at follow-up was similar across the three randomized groups; the percentage of missing interviews at any particular month ranged from 24% to 38%. Furthermore, all assessments were independent of attending clinical treatment sessions, so that we were able to obtain follow-up data on many women who dropped out of treatment. Among those who did not complete treatment, 71% completed three or more assessments; among those who did complete treatment, 92% completed three or more assessments. To impute missing data, we performed multiple imputation using an approximate Bayesian bootstrap method (Rubin, 1987) with hot-deck cells determined using predictive mean matching (Little, 1988). Outcomes were imputed one variable at a time chronologically. Each imputation model was conditioned on age, ethnicity, insurance status, marital status, education, amount of treatment received, the previous month outcome, and the subsequent month outcome. Imputations were done separately by treatment group. SAS statistical software version 8.2 was used for all analyses (SAS Institute Inc., 1999–2001).

Results

Care Received

Despite our active outreach, engagement, and facilitation efforts, many women did not complete treatment. Of those assigned to medication treatment (\( n = 88 \)), 67 (76.1%) received 9 or more weeks of a therapeutic dose of antidepressant medications, and 41 (47%) received the full 6 months of medication. Eighteen (20%) were switched to bupropion because the participant either did not tolerate paroxetine or did not respond (50% reduction in HDRS by 9 weeks). Overall, the mean number of medications sessions attended was 4.7 (\( \pm 3.37 \) SD), and the average number of weeks on medication was 19.2 (\( \pm 3.117 \) SD). Of the women assigned to psychotherapy (\( n = 90 \)), 32 (35.5%) received six or more CBT sessions. Fifteen (17%) were partial responders (reported less than a 50% reduction in Beck Depression Inventory (Beck, Steer, & Brown, 1996) scores given by clinicians during treatment) and were offered an additional 8 sessions of CBT. Among those completing a course of CBT, 15 completed group and 17 completed individual treatment. The average number of CBT sessions attended was 4.2 (\( \pm 4.05 \) SD), and for those who were offered an additional 8 sessions of CBT, the average number of additional sessions attended was 3.8 (\( \pm 2.82 \) SD). Of the women assigned to community referral (\( n = 89 \)), 74 (83.1%) failed to attend even one session. Only 8 women attended four or more sessions of community care.

Women received minimal care during the posttreatment period of follow-up. During the follow-up, women assigned to medications reported a mean of 0.64 (\( \pm 2.7 \) SD) mental health visits; women assigned to psychotherapy made 1.06 (\( \pm 3.6 \) SD) mental health visits; and women assigned to community referral made 1.36 (\( \pm 3.6 \) SD) mental health visits.
1-Year Intent-to-Treat Depression Outcomes

We examined changes in depression symptoms for individuals as a result of randomization to medications, CBT, or community referral (regardless of receipt of treatment). The overall multivariate F tests comparing the intercept, slope, and quadratic slope were positive for medication in comparison to community referral ($F = 11.6, p < .001$) as well as for CBT in comparison to community referral ($F = 3.3, p = .02$). Parameter estimates are presented in Table 2. As indicated by the significant linear slopes, depressive symptoms declined for all groups over time. The significant quadratic curves for both medications and CBT indicate a steeper initial decline in depression scores followed by a leveling out of scores compared with the community referral group, which did not produce as steep of an initial decline. These results are represented in Figure 2. Medication-assigned women also differed significantly from CBT-assigned women in the quadratic slope ($F = 4.03, p = .01$). As can be seen, the medication group showed the steepest initial decline in depression symptoms followed by the CBT group.

We examined differences in remission rates among treatment groups at Month 12 (defined as HDRS score $\leq 7$ and a 50% change from baseline to Month 12), using multiply imputed 12-month scores to include all participants as randomized. Results are given in Table 3. At 12 months, 50.9% of those assigned to antidepressants, 56.9% assigned to CBT, and 37.1% assigned to community referral were no longer clinically depressed. Odds ratio comparison indicates that this difference in rates of remission between those assigned to the medication group and those assigned to community referral nears significance ($p = .08$) and is significant comparing those assigned to CBT with those referred to community care ($p = .01$).

1-Year Intent-to-Treat Functional Outcomes

For social functioning, the multivariate F test comparing the intercept, slope, and quadratic slope of the medication group to those referred to community care was significant ($F = 4.2, p = .01$). The tests comparing CBT to the referred and the medication groups were not significant. Parameter estimates are presented in Table 2. The quadratic slope of the medication and CBT interventions were significant. As can be seen in Figure 3, initial sharp increases in social functioning occur for those assigned to medications, whereas those assigned to community referral show a slower, steady increase in functioning over time. The steeper increase in social functioning for the CBT participants was not statistically different from that of the referred group.

For impairment in instrumental role functioning (employee, homemaker, student), the multivariate F test comparing the intercept, slope, and quadratic slope of the medication group to the referred group was significant ($F = 5.14, p = .002$). The tests comparing CBT with medication and referral were not significant. Parameter estimates are reported in Table 2. The quadratic slope was significant only for medications. As can be seen in Figure 4, all participants' maladjustment in functioning in their major role decreased over time (as measured by the linear slopes), but those assigned to the medication condition reported steeper initial de-
clines compared with their CBT-assigned and community-referred counterparts.

Outcomes of Treatment Completers

Our intent-to-treat analyses represent likely outcomes of offering depression care, with active outreach, to low-income women seen in county entitlement services, thus offering policy-relevant information on what outcomes could be expected if these services were truly offered in these settings. In addition, we examined the subsample who actually received what we considered minimally adequate care compared with those who received no care to better understand the clinical benefit of treatment for this disadvantaged population. For this purpose, we compared those women in the medication group who received at least 9 weeks of guideline-concordant treatment (n = 67) and those in the psychotherapy group who received at least 6 CBT visits (n = 32) with those in the community referral group who did not receive any mental health treatment (n = 74). In these analyses, the CBT cell is small, with limited power for detecting differences. For the treatment completers, the results of the multivariate F test comparing the intercept, slope, and quadratic slope of the medication group versus the untreated group in depressive symptoms over 12 months was significant (F = 10.6, p < .01), as was the test comparing CBT with the untreated group (F = 3.65, p = .01). The completers among the medication and CBT groups did not differ significantly

Table 3
Odds of Remission and Percent Remitted From Depression as Defined by an HDRS Score of ≤7 and a 50% Decrease in Depressive Symptoms by Month 12

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Odds ratio (OR)</th>
<th>OR 95% CI</th>
<th>OR p value</th>
<th>% recovered</th>
<th>% recovered 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intent-to-treat sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral (n = 89)</td>
<td>ref</td>
<td></td>
<td></td>
<td>37.1</td>
<td>26.4–47.8</td>
</tr>
<tr>
<td>Medication (n = 88)</td>
<td>1.76</td>
<td>0.93–3.33</td>
<td>0.08</td>
<td>50.9</td>
<td>38.6–63.2</td>
</tr>
<tr>
<td>CBT (n = 90)</td>
<td>2.24</td>
<td>1.20–4.19</td>
<td>0.01</td>
<td>56.9</td>
<td>46.3–67.5</td>
</tr>
<tr>
<td>Treatment completers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untreated (n = 74)</td>
<td>ref</td>
<td></td>
<td></td>
<td>35.2</td>
<td>23.2–47.3</td>
</tr>
<tr>
<td>Medication (n = 69)</td>
<td>2.23</td>
<td>1.06–4.70</td>
<td>0.04</td>
<td>54.8</td>
<td>40.6–69.0</td>
</tr>
<tr>
<td>CBT (n = 32)</td>
<td>2.83</td>
<td>1.09–7.38</td>
<td>0.03</td>
<td>60.6</td>
<td>41.5–79.8</td>
</tr>
</tbody>
</table>

Note. Analysis was conducted on multiply imputed data. HDRS = Hamilton Diagnostic Rating Scale; OR = odds ratio; CI = confidence interval; CBT = cognitive–behavioral therapy.
from one another in depressive symptoms experienced over the 1-year follow-up.

As shown in Table 4 and Figure 5, significant quadratic slopes were found for both the medication and CBT groups. The medication and CBT interventions resulted in steep initial declines in depressive symptoms, with a leveling off over time. There was no significant linear or quadratic slope for the community-referred, untreated group. To examine rates of remission for the treatment completers (see Table 3), we used multiply imputed 12-month scores and found that 54.8% of
those completing antidepressant care, 60.6% of those completing CBT care, and 35.2% of those referred to but not entering antidepressant care and CBT groups had greater odds of achieving this outcome than did the untreated group (ps = .04 and .03, respectively).

For 12-month social functioning, the multivariate $F$ test comparing the intercept, slope, and quadratic slope of the treated medication group with the untreated group was significant ($F = 5.8$, $p < .001$), as was the test comparing CBT patients with referral patients ($F = 3.27$, $p = .02$). The test comparing medications with CBT was not significant. Parameter estimates are presented in Table 4. The linear slopes are significant for all groups, but the quadratic slope was significant for medication and cognitive therapy only. Results are presented in Figure 6. As can be seen, all participants improved in social functioning over time, but those completing medication and CBT treatments reported steeper initial increases in functioning compared with the untreated controls.

Finally, for 12-month role impairment, the multivariate $F$ test comparing the intercept, slope, and quadratic slope of the treated medication group with the untreated group was significant ($F = 4.3$, $p = .01$), whereas those comparing CBT with the untreated referral group and the medication group were not significant. Parameter estimates are presented in Table 4. The quadratic slope was significant only for medications. Results are presented in Figure 7. As can be seen, all participants reported decreases in impairment in their primary instrumental role over the year, but there was a steeper initial decline for those completing at least a minimal course of medications.

### Imputation of Results

To examine the robustness of our longitudinal findings to assumptions about missing data, we reexamined the major analyses using five imputed data sets. We replicated our finding that the medication intervention was superior to community referral in decreasing depressive symptoms in the intent-to-treat sample ($p < .001$) and the complete sample ($p < .001$), in improving social functioning (intent-to-treat: $p = .002$; completer: $p < .001$), and in improving instrumental role functioning (intent-to-treat: $p = .02$; completer: $p = .05$). We did not replicate the results for the CBT intervention except that CBT was superior to community referral in improving social functioning in the treatment completer sample ($p = .01$).

### Discussion

These results demonstrate the continued effectiveness over 1 year of short-term medication and psychotherapy interventions with CBT.

---

### Table 4

<table>
<thead>
<tr>
<th>Overall comparisons</th>
<th>Medication vs. untreated</th>
<th>CBT vs. untreated</th>
<th>Medication vs. CBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall comparisons</td>
<td>Hamilton Diagnostic Rating Scale*</td>
<td>Intercept</td>
<td>Slope</td>
</tr>
<tr>
<td>Medication vs. untreated</td>
<td>10.63 (3, 791)</td>
<td>&lt;.001</td>
<td>8.97</td>
</tr>
<tr>
<td>CBT vs. untreated</td>
<td>3.65 (3, 791)</td>
<td>.01</td>
<td>5.77</td>
</tr>
<tr>
<td>Medication vs. CBT</td>
<td>1.39 (3, 791)</td>
<td>.24</td>
<td>4.35</td>
</tr>
</tbody>
</table>

Note. Time was centered before entering the regression model to reduce collinearity between the linear and quadratic terms. Each regression model also controlled for ethnicity, baseline depression, and Baseline Depression x Time interaction. The coefficients for these terms are not shown.

---

1 As an additional test of the robustness of our findings, we reimpined the data using last observation carried forward. We replicate our major findings in these analyses.
for low-income and minority women suffering from depression. We find that women who were not seeking depression treatment, but who screened positive for depression in county entitlement programs, report rapid reduction of depression symptoms and return to functioning when offered medications or psychotherapy, along with the educational outreach and support, such as child care and transportation, necessary to obtain care. Policies supporting implementation of depression screening and care for women seen in public entitlement programs would decrease suffering and increase functioning among poor and minority women.

Our results also suggest that low-income and minority women who complete even minimally adequate treatment report clinical improvement similar to more advantaged populations. More than half of those who completed 9 weeks or more of an adequate dose of paroxetine or 6 weeks or more of CBT were recovered at 1 year. Only 35% of those obtaining no care had recovered. These remission rates for our treated population are relatively similar to those reported by Hollon et al. (2005), who examine outcomes of patients who responded to CBT or medications and were monitored for 12 months. Over the course of 12 months following response to care, 69.2% exposed to CBT, 52.8% continued on antidepressant medications, and 23.8% withdrawn from antidepressant medications were recovered. Providing care to disadvantaged women results in long-term outcomes quite similar to more advantaged groups.

However, the disadvantaged women in our trial, who were not seeking depression care initially, were much less likely to complete care compared with those in efficacy trials of more advantaged populations who were seeking care for their depression. For example, DeRubeis et al. (2005) found that 89% of their predominantly White and educated participants assigned to medications and 85% assigned to cognitive therapy remained in care at 8 weeks. As reported previously, in our disadvantaged sample, only 67% received 9 or more weeks of medication and 35% completed six or more sessions of CBT.

Overall, the medication intervention resulted in stronger response compared with CBT when pooling across the 12 months. Over the course of the 12 months, participants receiving medication had significantly fewer depressive symptoms than those assigned to CBT. This result is likely influenced by the minimal number of psychotherapy treatment sessions we were able to get the women to attend. The CBT intervention had rates of remission at 12 months equal to that of the medication condition. Thus, even minimal exposure to CBT might be useful for disadvantaged women over time. This suggests the importance of developing very brief, structured CBT interventions that might be introduced to women in the course of their county health care and may provide them with useful strategies for improving depression outcomes.

Medications resulted in advantages over both CBT and community referral in functional outcomes. The CBT was less effective in improving functional outcomes. Again, this may in part be due to the low number of sessions the women attended. Those who completed six sessions of CBT did report improved social functioning but not improved role functioning. Overall, the functional outcomes appear to mirror our symptom outcomes. As women respond to depression care, their overall functioning appears to improve.
We did not find differences in response to care between African American and Latina women, the samples we had adequate power to examine. In a review of the literature on outcomes of care for ethnic minorities (Miranda et al., 2005), African American and Latino patients were found to respond to depression interventions that are evidence based similarly to White participants. An investigation of depressed, impoverished women in Mexico (Lara, Navarro, Rubi, & Mondragón, 2003) found that a 6-week group intervention tailored to the problems and issues of these women was no more effective in treating their depression than one brief psychoeducational meeting. It seems likely that culturally tailored care alone is ineffective. Alternatively, culturally sensitive outreach and engagement seems essential to bringing low-income and minority women into evidence-based care.

In this study, active, culturally sensitive education and encouragement were needed to get women into care. Even so, we found that engaging low-income women into care who were not initially seeking mental health treatment was particularly difficult. For example, despite outreach and help with transportation and babysitting, only 36% of the women assigned to CBT attended six or more therapy sessions. It is possible that the care offered in this study was not experienced as culturally appropriate for these low-income women. Furthermore, the very demanding nature of these women’s life (multiple jobs, young children) often precluded consistent attendance at care. Despite this problem, the women appear to have benefited greatly from the care over a year’s follow-up. In the future, efforts to embed care for depression within settings that these women attend frequently may be more effective than delivering them in the family planning and entitlement program settings of this study. For example, offering psychotherapy in Head Start programs as women pick up their children might be a cost-effective and efficient setting to provide such care. Churches may also be effective settings to provide access to care for these women. Clearly, determining ways to facilitate women attending care for depression is important for this disadvantaged population.

Several limitations to this study should be noted. First, all measures were self-report, although the study measures have been used extensively in previous depression studies. Second, nearly half of the women who screened positive for depression did not follow up with diagnostic interviews, and others claimed to no longer be depressed when interviewed later. Thus, the extent to which this sample represents depressed low-income young women is not known. Third, women were queried by telephone multiple times for study assessments; the impact of these multiple contacts is not known. Fourth, although the psychotherapists were closely supervised, we did not measure adherence to the manualized treatment nor did we measure competence of the therapists. Finally, the impact of psychotherapy or medication apart from the enhancements necessary to bring women to treatment, including providing babysitting, transportation, and up to four educational sessions conducted by the clinical providers, cannot be determined in this study.

We conclude that providing evidence-based antidepressants or psychotherapy care for low-income women decreases their suffering. Antidepressant medications result in rapid decreases in de-
pressive symptoms, and both treatments achieve high rates of remission by 12 months.

References


Received September 7, 2004
Revision received September 22, 2005
Accepted September 29, 2005

---

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at [http://watson.apa.org/notify](http://watson.apa.org/notify) and you will be notified by e-mail when issues of interest to you become available!
The complexity of gender differences in health (i.e., men's lower life expectancy and women's greater morbidity) extends beyond notions of either social or biological disadvantage. Gaps remain in understanding the antecedents of such differences and the issues this paradox raises regarding the connections between social and biological processes. Our goals in this analytic essay are to make the case that gender differences in health matter and that understanding these differences requires an explanation of why rational people are not effective in making health a priority in their everyday lives. We describe some salient gender health differences in cardiovascular disease, immune function and disorders, and depression and indicate why neither social nor biological perspectives alone are sufficient to account for them. We consider the limitations of current models of socioeconomic and racial/ethnic health disparities to explain the puzzling gender differences in health. Finally, we discuss constrained choice, a key issue that is missing in the current understanding of these gender differences, and call on the social science community to work with biomedical researchers on the interdisciplinary work required to address the paradoxical differences in men's and women's health.

**Issue Section:** TOPIC 2. THE INTERFACE OF BIOLOGICAL AND EXPERIENTIAL CONDITIONS IN HEALTH INEQUALITIES

IT has long been established that women live longer than men, yet they have higher morbidity rates. Men experience more life-threatening chronic diseases and die younger, whereas women live longer but have more nonfatal acute and chronic conditions and disability. Furthermore, although men's and women's overall rate of serious mental illness are similar, the most common mental health disorders differ by gender. There is considerable evidence that both biological systems and social processes underlie these perplexing patterns, but what is missing in our understanding of these gender differences is an explanation of how the social and biological factors combine to produce these paradoxical differences in health for men and women.

Our primary goals in this article are to make the case that gender differences in health matter and that understanding these differences requires an explanation of why rational people are not effective in making health a priority in their everyday lives. We describe some salient gender health differences and indicate why neither social nor biological perspectives alone are sufficient to account for them, then we consider the limitations of current models of racial/ethnic and socioeconomic health disparities to fully explain gender differences. Last, we discuss a key issue, constrained choice, that is missing in the current understanding of gender differences in health, and we call on the social science community to work with biomedical researchers on the interdisciplinary research that is required to address the paradoxical differences in men's and women's health.

**THE GENDER AND HEALTH PARADOX**

As noted above, differences in men's and women's physical health are paradoxical. Women live longer than men, yet they have higher morbidity rates and in later years a diminished quality of life. In fact, although the size of the gender gap and pattern of longevity varies considerably by country, women outlive men in every region and almost every country of the world (United Nations, 2000). Although many reasons for the variation have been identified, biological theories are not considered a sufficient explanation for the cross-national differences.
Since the turn of the 20th century, American women's life expectancy has exceeded men's, with women experiencing lower mortality rates in every age group and for most of the age-adjusted leading causes of death. Although life expectancy has been increasing for both men and women, the gender gap in the United States, currently 5.4 years, has been closing since 1980 when men's gains began to exceed women's owing in large part to men's decreasing mortality from cardiovascular disease (CVD) and cancer (Centers for Disease Control and Prevention [CDC], 2000; National Center for Health Statistics [NCHS], 2003).

The paradox of men's higher mortality and lower morbidity compared with women can be explained by gender differences in age-related patterns of disease, including women's increased risk for CVD after menopause (Verbrugge & Wingard, 1987). Although the three leading age-adjusted causes of death are the same for men and women (heart disease, cancer, and stroke), men have more life-threatening chronic diseases at younger ages, including coronary heart disease, cancer, cerebrovascular disease, emphysema, cirrhosis of the liver, kidney disease, and atherosclerosis. In contrast, women face higher rates of chronic debilitating disorders such as autoimmune diseases and rheumatologic disorders as well as less life-threatening diseases such as anemia, thyroid conditions, gall bladder conditions, migraines, arthritis, and eczema. Women also have more acute conditions such as upper respiratory infections, gastroenteritis, and other short-term infectious diseases (NCHS, 2003).

In contrast to the paradoxical gender differences in physical health, men's and women's overall mental health is similar, but specific conditions differ by gender. Women experience substantially higher rates of depression and anxiety disorders than men, whereas men have higher rates of substance abuse, antisocial behavior, and suicide (Kessler et al., 2003a,b).

Below we examine three reference conditions—CVD, immune function and disorders, and depression—that exhibit substantial gender differences. Although our examination is not exhaustive, it paints a more complex portrait of specific patterns of gender differences in health than previously articulated. Later we will consider some of the primary social and biological explanations of these patterns.

**Cardiovascular Disease**

Even though the prevalence and age-adjusted death rate of CVD are greater in men than in women, more women ultimately die of the disease because of their greater life expectancy, the older age at onset, and the range of CVD risk factors associated with aging. For example, in terms of total deaths since 1984, CVD has claimed the lives of more females than males (NCHS, 2003). Crimmins and colleagues (2002) provide interesting data on life expectancy by estimating years lived with and without disease by birth cohorts of men and women over the life course. Because more women survive to older ages, the gender difference in years lived with disease is even greater than years of life expectancy. For example, a cohort of women will experience 70% more years of life after age 65 with hypertension than a similar sized birth cohort of men. Likewise, a group of older women also will spend more years with CVD than will older men. Thus, the modal patient undergoing treatment for CVD and hypertension is likely to be a woman beyond middle age.

Throughout most of the last two decades, clinical researchers focused on explaining the earlier onset of CVD in men and the biological mechanisms that contribute to men's lower average in life expectancy. This research emphasis, reflected in large single-sex longitudinal clinical trials directed at prevention of CVD in men, contributed over time to significant reductions in male mortality from CVD and helped to narrow the gender gap in longevity. While the life-expectancy gender gap is narrowing, the CVD mortality gap continues to widen (NCHS, 2003).

A confluence of factors in recent decades, including the women's health movement, led to substantial increases in women's inclusion in research and in turn to dramatic changes in knowledge and understanding regarding women's risk of CVD, but little insight into the antecedents of gender differences in risk. In 1991, the Women's Health Initiative (WHI), the largest preventive study of its kind in the United States, was launched to address this gender gap. Although the WHI represented a huge step forward in gathering scientific evidence regarding women's health, it did not, for example, question the basic physiological explanations of role of hormones in women's extensive CVD mortality. For example, the study did not draw on a broader interdisciplinary model and instead was designed to test prevailing biomedical
explanations, based on prior hypothesis and findings primarily from observational studies that indicated protective effects of estrogen and diet in preventing the development of CVD.

Ironically, the primary clinical trial arm of the planned 15-year hormone replacement therapy (HRT) intervention was halted in 2002 after a mean of 5.2 years of follow-up. The data and safety and monitoring board recommended stopping the trial because HRT failed to protect against CVD and women receiving the estrogen/progesterone combination were also found to have an increased risk of invasive breast cancer and other negative health effects (Fletcher & Colditz, 2002; Rossouw et al., 2002). Although the lessons that will be learned ultimately from these developments are yet to be determined, it is clear that investigations involving only women—or only men—are still valuable. However, as such, single-sex studies cannot address the larger issue of gender differences.

Immune Function and Disorders

Researchers and clinicians are challenged and perplexed by the sex-linked patterns of immune function and disease. Women have a greater risk of autoimmune disorders and a higher risk of genetic immune suppression disorders than men; for example, the female/male ratio for autoimmune thyroid disorders is 15:1, systemic lupus 9:1, and rheumatoid arthritis 3:1 (Office for Research on Women's Health, 1992). Much of the disability women experience from rheumatologic and thyroid disorders especially in middle and older ages is attributable to autoimmune disease (CDC, 2000).

Females of most, if not all, species generate a more robust immune response after challenge with an infectious agent and respond to natural and vaccine exposures by producing substantially more antibodies than males. Women's different and somewhat more intense immune response provides them with higher levels of passive immunity during pregnancy and the ability to pass on a substantial level of protective antibodies to infants during breast feeding (Grossman, Roselle, & Mendenhall, 1991; Hegde, 1991). But, pregnancy exacerbates the course of some autoimmune diseases (e.g., lupus), while it improves the course of others (e.g., rheumatoid arthritis). In other words, women's more robust immune systems appear to put them at greater risk of disorders in which their own immune system attacks their bodies.

Sex hormones also tend to modulate the immune response: What biologists refer to as distinct immune hormonal environments contribute to some of the sex differences in the prevalence of autoimmune disorders. In addition, men and women tend to develop different autoimmune conditions. While the origin of these differences appears to be primarily biological, some physicians believe that gender, and in particular the stress process (Pearlin, 1989), also affect autoimmune disease incidence and severity (Legato, 2002; Lockshin, Gabriel, Zakeri, & Lockshin, 1999). A rapidly developing research literature suggests that social/emotional factors combine with biological processes to affect health (see, e.g., Cacioppo et al., 2002; Kiecolt-Glaser, McGuire, Robles, & Glaser, 2002a; Seeman & McEwen, 1996; Seeman, Singer, Rowe, Horwitz, & McEwen, 1997).

Various possible pathways have been described through which psychological factors impact on immune function (Kiecolt-Glaser, McGuire, Robles, & Glaser, 2002b). For instance, positive relationships seem to improve function, while discordant relationships seem to have a negative impact. Even though acute and chronic stressors may operate through different physiologic pathways, coping resources such as social support mediate the impact of both (Baron, Cutrona, Hicklin, Russell, & Lubaroff, 1990; Cohen & Herbert, 1996). Although there is some debate about the type and number of social ties by gender and their impact on men's and women's health, new work on gender and stress reactivity suggests that women may actually respond to stress by “tending and befriending” rather than men's tendency to “fight or flight” (Taylor, Klein, Lewis, Gruenewald, Gurung, & Updegraff, 2000).

Depression

The World Health Organization (WHO) ranks major depression as one of the most burdensome diseases in the world (WHO, 2002). Moreover, a growing body of research links depression with physical health, further illustrating the complexity of assessing the antecedents of gender differences in health.
Women's rates of depressive disorders are between 50% and 100% greater than men's (Kessler et al., 2003a; Mirowsky & Ross, 2003). These differences occur in both treated and community samples (Nolen-Hoeksema, 1987, 1990). The National Comorbidity Survey Replication substantiated women's higher lifetime rates of major depressive disorder (Kessler et al., 2003a,b). It is now well established that women's higher rates of depression reflect a real gender difference in health rather than an artifact of help-seeking behavior or willingness to report symptoms (Mirowsky & Ross, 1995; Nazroo, Edwards, & Brown, 1998).

Until the recent men's health movement, women's disproportionate depression rates generated the erroneous impression that men were comparatively immune to depression (Courtenay, 2000a,b). The underdiagnosis of depression in men has been attributed to clinicians' failure to recognize symptoms, to men's unwillingness to seek help for such feelings, and to their tendency to cope with the feelings through drinking, drug use, and other private activities or actions (Chino & Funabiki, 1984; Nolen-Hoeksema, 1987, 1990).

The gender difference in the prevalence of depression occurs at all ages. The risk for depressive disorders begins to rise in midpuberty and persists through the mid-50s; however, the gender gap appears to be the greatest during reproductive years (Bebbington, 1996; Piccinelli & Wilkinson, 2000). There is some controversy about the determinants of gender differences and discrepancy between cross-sectional and longitudinal findings regarding the course of depressive disorders. Although men and women differ in the age and rates of onset (young males have higher rates until early adolescence), cross-sectional studies indicate that once major depression develops, the course is similar for both genders. The continuing debate about the course of depressive illness in women compared with men does not alter the consistently higher lifetime prevalence rates for women or the fact that depressed women also are more likely than are men to have comorbid anxiety, while men are more likely to have comorbid substance abuse or dependence (Endicott, 1998; Gregory & Endicott, 1999; Kessler et al., 2003a). A recent review of research on gender differences in depression concluded that while there are some psychosocial risk factors contributing to women's higher rates, the determinants of these differences remain unclear because models combining developmental physiological and psychosocial processes are lacking (Piccinelli & Wilkinson, 2000).

**PREVAILING EXPLANATIONS FOR THE PARADOXICAL GENDER DIFFERENCES IN HEALTH**

**Biological Explanations**

Many of the biological explanations of women's greater longevity emphasize the health advantages that accrue from different hormones and physiological systems that facilitate pregnancy and childbirth—advantages that have long been hypothesized to lower women's risk of coronary heart disease prior to menopause. However, results from several recent clinical trials of HRT are challenging that assumption (Fletcher & Colditz, 2002; Liddle, 2003). In addition, the hypothesized advantages are not consistent with women's higher morbidity compared with men.

One could easily imagine that evolutionary selection factors would produce strong women whose ability to bear children also provided advantages in terms of both morbidity and mortality. Such evolutionary arguments build from the premise that natural selection favors individuals that produce more offspring, as their genes are more highly represented in subsequent generations of members of a species. The theory of natural selection and its impact on the prevalence and distribution of specific genes are at the core of biology. Although the evidence supporting explanations of the effects of natural selection on survival through reproductive ages is strong, the evidence is much weaker when it comes to explaining differences in men's and women's longevity well beyond prime reproductive age. Purely biological explanations, including the strong pursuit to identify the genetic basis of disease, still fall short of fully explaining the gender paradox of morbidity and mortality.

However, numerous interdisciplinary biomedical studies indicate that there are interactions between the cardiovascular system, immune functioning, and psychological processes (see, e.g., Hemingway & Marmot, 1999; Kiecolt-Glaser et al., 2002a; Lovallo, 1997). For instance, considerable evidence links clinically diagnosed major depression with increased mortality in general and CVD in particular. Although these
studies do not show that depression is a predictor of mortality per se, symptoms of depression are associated with poor health and functional status, as well as increased disability, health care utilization, and cost of health services. A 13-year prospective study showed 4.5 times greater risk of a heart attack among those with major depression (Pratt, Ford, Crum, Armenian, Gallo, & Eaton, 1996). In another study, mortality was four times higher among depressed heart attack patients (Frasure-Smith, Lesperance, & Talajic, 1995). A prospective study of a large cohort of White women aged 67 years or older found that depressive symptoms were associated with increased rates of cardiovascular mortality (Whooley & Browner, 1998).

Depression is also associated with immune function and disease severity by magnifying pain and disability (Staats, 1999). Pain and depression can amplify each other: Pain increases heart rate and blood pressure; it enhances secretion of stress-related hormones including catecholamines and cortisol; and it deregulates a range of immunological activities (Kiecolt-Glaser, Page, Marucha, MacCallum, & Glaser, 1998). Thus, gender differences in mental health may contribute in unknown ways to gendered patterns of physical health (McDonough & Walters, 2001) and vice versa.

Social Explanations

Purely social explanations of gender differences in health emphasize men's and women's social position and their differential access to protective resources (including income, education, safe parks, and other areas in which to exercise) as well as their exposure to a range of environmental factors that negatively affect health (including exposure to toxins and other social and behavioral risk factors such as domestic violence, crime, smoking, poor diet). Social position mediates access to the positive and negative social and environmental factors that occur at the individual, household, community, and society levels. For example, gender differences in income may result from a variety of factors over the life course including social role-related expectations and activities, occupational choice, opportunities for employment and advancement, work-related skills and experience, as well as access to job benefits including health insurance, pensions, and other retirement income. Women are also more likely to be single parents and thus have a lower income per capita. Although the impact of income on health would be expected to be the same for men and women, the levels of income may vary by gender for a variety of reasons other than the job one currently has or pay inequity. For instance, welfare and income entitlements such as Social Security are, in fact, tied to both one's own employment and income history and that of one's spouse.

Other risk and protective factors may differ by gender in both exposure and their impact on health. For example, toxin exposure may be related to patterns of employment (e.g., asbestos exposure leading to mesothelioma) and social roles (e.g., industrial or household use of pesticides or other chemicals or to exposure to second-hand smoke). Furthermore, exposure to a specific toxin may result either in somewhat different rates of the same problems (e.g., lung cancer) or in different health problems for men and women (e.g., breast or prostate cancer). Moreover, other resources that vary on average by gender such as education may provide additional knowledge, opportunity, and income with which to avoid multiple risk factors. Clearly, the distribution and use of toxins are regulated by social policy, but so, too, is education. Thus, in addition to being an individual resource, education is also a product of both governmental policies and community choices and priorities.

Sociological theory offers considerable insight into women's greater morbidity and the gendered patterns of psychological disorders. But for the most part, sociologists and other social scientists have not fully examined the diverse antecedents of gender differences in mortality or the types and prevalence of diseases. Instead, the research on social determinants has concentrated on socioeconomic position, a "fundamental social cause" as theorized by Link and Phelan (1995), focusing on a causal chain whereby unequal distribution of social and economic resources and related variation in exposure to stress lead to health disparities (Berkman & Kawachi, 2000; House & Williams, 2000; Phelan, Link, Diex-Roux, Kawachi, & Levin, 2004). Moreover, applying the implicit logic of disadvantage, a majority of mental health studies have concentrated on explaining women's higher levels of depression, but not men's higher levels of substance abuse and antisocial behaviors. Thus, sociological explanations of gender differences in health examine variation in gender roles and identities, the ways in which racial and socioeconomic inequality contributes to differences in men's and women's exposure to stressors, the meaning and impact of particular stressful events and coping resources, and the physical and psychological consequences of each of these factors.
DO MODELS OF RACIAL AND SOCIOECONOMIC DISPARITIES ADEQUATELY EXPLAIN THE GENDER HEALTH PARADOX?

Clearly, racial and SES (socioeconomic status) disparities overlay the gender differences in health. Although gender differences in morbidity and mortality vary across racial/ethnic and socioeconomic groups, they also persist within each of these groups (NCHS, 1998; Rieker & Bird, 2000; Williams, 2002, 2003). Moreover, explanations of racial and socioeconomic disparities focus in large part on the multiple and cumulative consequences of poverty and discrimination, among other factors (Brunner & Marmot, 1999; House & Williams, 2000; Phelan et al., 2004; Smaje, 2000). These factors are critical to understanding racial/ethnic and socioeconomic variation in morbidity and mortality among men and among women, but they cannot fully explain women's greater longevity and men's earlier deaths and only partially explain women's pattern of greater morbidity relative to men.

Unlike racial/ethnic and socioeconomic disparities in health, men's and women's social and economic needs can and do differ in part because of biological and longevity differences in their lives. For example, because women are more likely to become single parents, caregiver to an aging spouse, widowed, and to live well into their 80s and beyond, they also have different social and economic needs and thus may not be served equally well by social policies designed to ignore these differences. In fact, a wide range of social and economic policies, including but not limited to those directly aimed at public health and health care, can create or exacerbate differences in health status and behaviors between gender groups as they do with racial and economic groups (see, e.g., Graham, 2004).

WHAT IS MISSING IN OUR UNDERSTANDING OF GENDER DIFFERENCES IN HEALTH?

A growing body of research indicates that the complexity of the gender health differences extends beyond narrow concepts of the relative disadvantage or advantage of men's and women's biology or the social organization of their lives. Consequently, we need a model of men's and women's health that takes into account factors other than inequality of resources, discrimination, and other unfair treatment, to understand what aspects of the broader array of differences in men's and women's lives contribute to the morbidity/mortality paradox.

Treating gender as a fundamental cause sheds little light on how social factors contribute to the gender health paradox, including whether and how they affect or interact with biological processes. We cannot simply substitute gender for race or socioeconomic status in existing models of health disparities because they are not constructed to capture the complex way in which men and women are advantaged and disadvantaged in both mortality and morbidity. Such models are based on the implicit hypothesis that the health effects produced by inequality and discrimination go in the same direction for biological and social factors. While most models of social determinants of health incorporate biological processes, only recently have researchers begun to simultaneously assess these interdisciplinary factors (see e.g., Sen, George, & Ostlin, 2002). The issue in the case of racial/ethnic and socioeconomic disparities is how to parcel out the social and biological antecedents, assuming that they do not confound each other, as may well occur with gender differences in either morbidity or mortality. For example, men's general economic advantages would not on the face of it be likely to explain their earlier mortality compared to women (although socioeconomic status might partially explain the mortality difference between White and African American men). If social factors contribute to gender differences in ways that extend beyond inequality, models of gender differences need to capture the specific processes and pathways that lead to health outcomes. Although researchers have examined some of the pathways through which social factors might differentially affect men's and women's health (including positive and negative health behaviors, exposure to stressors, and coping styles), for the most part, this approach has not been used to explain gender differences in health.

There is no comprehensive social framework for considering the influence of multiple levels of social factors, including individual agency or choice, on gender differences in health. Such a framework also needs to incorporate contextual effects at the levels of family, community, and
social policy in ways that extend beyond models of gender inequality and inequity. This approach would also facilitate the broader integration of social and biological processes and contribute to a better understanding of differences in men's and women's health.

**WHAT IS NEEDED TO GAIN A BETTER UNDERSTANDING OF GENDER DIFFERENCES IN HEALTH?**

**Toward a Sociology of Constrained Choice**

Nearly two decades ago, Verbrugge and Wingard (1987) pointed out that neither a strictly biomedical interpretation of the data nor one based solely on social factors adequately explains why the health of men and women differs in such obvious ways. Considerable evidence suggests that both biological systems and social processes underlie these perplexing patterns.

An explanation of how such factors combine to produce the paradoxical gender differences is still missing from our understanding of men's and women's health. We call on social scientists and biomedical researchers to work together to meet the challenge of developing models that capture the ways in which the gender health paradox differs from socioeconomic and racial/ethnic health disparities. To encourage this new direction, we are developing a model of health determinants that includes the concept we refer to as constrained choice, which we elaborate in chapter 2 of our book, *Gender and Health: The Effects of Constrained Choices and Social Policies* (Bird & Rieker, in preparation). This concept builds on the argument that Boudon (2003) and Light and Hughes (2001) make, distinguishing sociology of choice from economic models of rational choice theory.

Not all of the constraints that affect everyday choices and opportunities to make health a priority are the same for men and women. We contend that to explain gender differences, models of health determinants need to be modified to include these constrained choices. We are not posing constrained choice as a deviation from an idealized notion of “free” choice; rather, we are describing a process by which choices are made. We draw on the prevailing public health understanding of health disparities, which emphasizes the role of personal choices and health behaviors in enhancing or diminishing an individual's ability to live a long and healthy life.

At the individual level, these choices can influence and be influenced by biological processes (in a feedback loop). For example, constrained choices impact on men's and women's stress levels as they experience competing demands on their time and other resources, which can in turn affect their psychological and physical responses to stress. Such effects could occur through gender differences in exposure to role-related chronic stressors, which can ultimately exacerbate stress reactivity, leading to dysregulation and dysfunction in both cardiovascular and immune responses. Constrained choices may also impact positive and negative health behaviors and coping styles that affect both psychological and physical functioning. Consequently, we contend that gender differences in the constraints contribute to health disparities both directly and indirectly by affecting both men's and women's choices and their cumulative biological risk. An emerging body of work examines social and biological mechanisms (e.g., health behaviors and physiological pathways) through which psychosocial factors such as social integration, marital status, and social support affect health outcomes (for reviews of this literature, see Cacioppo et al., 2002). Some of these studies provide evidence of gender differences in pathways and in outcomes (e.g., Karlamangla, Singer, Greendale & Seeman, 2005; Loucks, Berkman, Gruenewald & Seeman, in press; Koivumaa-Honkanen, Honkanen, Viinamäki, Heikkilä, Kaprio, & Koskenvuo, 2000; Seeman, McEwen, Singer, Albert, & Rowe, 1997; Seeman et al., 2004; Seeman, Singer, Ryff, & Levy-Storms, 2002). However, research has yet to assess the contribution of constrained choice to these gender differences.

We intend the schematic framework depicted in Figure 1 for heuristic purposes and to address two central questions in this complex scenario of gender and health differences: (1) What keeps men and women from making health an everyday priority? (2) Are men and women acting rationally when they behave as if health is not a priority? Basically we argue that men's and women's opportunities and choices are to a certain extent constrained by decisions and actions taken by families, communities, and governmental policies. Although many of the constraints and their consequences for individual choice are similar for men and women, the health impact will vary somewhat owing to gender differences in biology and life experiences.
Figure 1 depicts the various contexts that influence men's and women's lives, linking decisions at each of these levels, which collectively impinge on the opportunities of individuals and groups to choose health. These connections between broader social contexts and individual choices are rarely transparent, and their health consequences are often underestimated and frequently overlooked. Below, we briefly elaborate on the impact of governmental policies on men's and women's health because the ways in which families and communities differentially affect men's and women's health have been explored to some extent.

Some examples of constrained choice include the impact of governmental and workplace policies on men's and women's health. Although a given decision or policy at any level does not necessarily aim to constrain choices in ways that affect health or to impact men and women differently, many do so. For instance, workplace policies that provide on-site childcare affect childcare decisions and stress levels for both men and women as they experience competing demands from work and parental roles. For mothers, it also affects the related decision of whether or how long to breastfeed (Jacknowitz, 2004), which can in turn affect women's hormone levels, their parenting, and ultimately their risk of breast cancer. A more obvious example, Social Security is primarily intended as a safety net to keep older adults out of poverty. Because of women's lower wages and greater longevity, they pay less into Social Security. Subsequently, they receive lower payments over a substantially longer average life span than that of men, and these lower payments can constrain their choices for a good diet, a safe place to live, and the ability to meet other basic needs that have health consequences. These examples highlight some of the ways in which social constraints can differentially affect men's and women's health.

Conclusion

A gap still exists in our understanding of the antecedents of gender differences in health and around the issues this paradox raises about the multifaceted connections between social and biological processes. For the most part, studies have failed to assess gender and to explain why rational people are not effectively making health a priority in their everyday lives. Existing influential models of social determinants of health and disparities are not sufficient to explain why the health of men and women differs in such puzzling ways. Moreover, a growing body of research indicates that the complexity of these paradoxical health differences extends beyond singular notions of the disadvantage or advantage of either men's or women's biology or socially organized lives.

Neither biological nor social research alone can answer the complex questions regarding the antecedents of the gender differences. Only a synthesis of these perspectives can move forward the much-needed interdisciplinary dialogues and investigations to close the knowledge gap. Although many significant voices are calling for such a synthesis to explain a variety of other health disparities, and new multidisciplinary fields are grappling with identifying and understanding the connections among biological processes and social factors, addressing gender differences is still overlooked. Perhaps the most obvious explanation of why they are overlooked is that the social and biomedical sciences have been working on the same gender paradox but from the confines of their own disciplines. Consequently, researchers in these diverse fields have not necessarily been part of the same discussion or even attempted to answer the same questions, much less tried to explain the larger paradox.

To encourage the social science community to think in new ways about this important issue, we are developing a model of constrained choice that addresses what we feel is missing in explanations of gender differences in health. In doing so, we seek to provide a focal point for cross-disciplinary research and to underscore the need for social scientists to collaborate with researchers in the biological sciences in the design of studies that examine the multifaceted relationships involved in gender health differences. Together, we are uniquely prepared to bring in a much-needed focus on the contribution of multiple levels of contextual effects and to advance interdisciplinary investigations.

We understand the difficulty of designing and obtaining funding for such studies. Yet without this collaborative interdisciplinary work, we will never reach a clear understanding of the paradoxical health differences between men and women nor gain significant new insight into keeping both men and women healthy (Bird & Rieker, 1999; Rieker & Bird, 2000).
Conceptualization of constrained choice

This article is adapted from chapters for a book in progress by the authors (Gender matters: The health effects of constrained choice and social policy) funded by a grant from NIH/National Library of Medicine.

We especially thank Stephanie Taylor, PhD, and Elizabeth Maggio for their insightful contributions to earlier drafts.

References


Treatment Seeking and Utilization among Women with Substance Use Disorders

Christine E. Grella, PhD

Since the 1970s the field of substance abuse treatment has witnessed enormous changes in the number of treatment providers, forms of treatment financing, and types of treatment modalities available. There has been increasing emphasis on developing, implementing, and disseminating evidence-based treatment for substance use disorders within community-based settings. Moreover, increasingly sophisticated epidemiological surveys have provided a more detailed understanding of the prevalence of substance use within the general population, the characteristics of those who have substance use disorders (SUDs), and the rates at which they receive treatment (Degenhardt, Chiu, Sampson, Kessler, & Anthony, 2007). Further, data collected from national surveys of treatment providers yield a composite picture of the characteristics of individuals participating in treatment and enable comparisons both across time periods and by types of settings in which individuals receive treatment.

These various population surveys and provider surveys, along with clinical research from individual and multisite studies, allow for an in-depth examination of the role of gender in substance abuse treatment participation. The goal of this chapter is to provide a synthesis of research regarding treatment seeking and utilization among women with SUDs, and the implications of this body of research for future research and policy. This chapter examines the influence of gender in the following areas: (1) the prevalence of SUDs and need for treatment in order to provide a context for understanding treatment utilization; (2) the barriers and facilitators of treatment utilization, including the complex and unique impact that being a parent has on treatment participation among women; (3) the pathways into substance abuse treatment through other service delivery systems with which women come into contact; and (4) interventions to increase treatment participation, including coerced treatment and motivational enhancements.

SEX DIFFERENCES IN PREVALENCE OF SUDs AND HELP SEEKING

Data from national prevalence surveys show that a greater proportion of men in the general population have a history of SUDs (both abuse and dependence); however, the sex differ-
ence is less pronounced with regard to drug use disorders (Kandel, 2000). For example, in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), males are 3.1 times more likely than females to have a lifetime alcohol use disorder, 2.3 times more likely to have alcohol abuse, and 2.6 times more likely to have alcohol dependence (Hasin, Stinson, Ogburn, & Grant, 2007). The differentials are less pronounced with regard to lifetime drug use disorders; males are 2.3 times more likely than females to have a lifetime drug use disorder, 2.2 times more likely to abuse drugs, and 1.9 times more likely to be drug dependent (Compton, Thomas, Stinson, & Grant, 2007). The gap between males and females is smallest with regard to the prevalence of past-year amphetamine use disorders; males are about 1.6 times more likely than females to manifest this disorder. In addition, women are more likely than men to engage in nonmedical use of prescription drugs, particularly narcotic analgesics and tranquilizers; women are twice as likely as men to report past-year abuse of these substances (Simoni-Wastila & Strickler, 2004).

Data from the National Survey on Drug Use and Health, conducted from 2004 to 2006, indicate that an annual average of 6.3 million women (9.4%), ages 18–49, needed treatment for a substance use problem (Office of Applied Studies [OAS]; 2007). Of the women ages 18–49 who met criteria for needing substance use treatment in the past year, 84.2% neither received it nor perceived the need for it. Only 5.5% of women in this age group had a perceived unmet treatment need. The survey further elicited reasons for not receiving substance use treatment among the women with an unmet treatment need: 36.1% stated they were not ready to stop using alcohol or illicit drugs; 34.4% could not cover their treatment costs because of no, or inadequate, health insurance coverage; and 28.9% did not seek substance use treatment because of social stigma.

National survey data have shown that help seeking for substance use problems among the general population is generally low, although it is higher among individuals who have multiple types of substance use problems, as well as co-occurring mental disorders (Wu, Ringwalt, & Williams, 2003). Yet, when controlling for number of disorders and other demographic characteristics, males with at least 1 past-year disorder were approximately 1.7 times more likely than females to report having received substance abuse services. Upon reviewing data from several national surveys, Greenfield and colleagues (2007) concluded that the proportion of women who participate in substance abuse treatment is lower than the population prevalence of SUDs in women relative to men. Hence, although most women with SUDs in the general population do not acknowledge a need for treatment, the rate of unmet treatment need may be greater among women than men.

Despite the high rates of unrecognized need for treatment among women, it is well established that women demonstrate an accelerated progression from initiation of substance use (particularly cannabis, alcohol, and cocaine) to onset of dependence and first admission to treatment, as compared with men (Hernandez-Avilia, Rounsaville, & Kranzler, 2004; Hser, Anglin, & Booth, 1987; Wagner & Anthony, 2007). When women do enter treatment, they typically report a shorter duration of use but more severe disorders (usually measured by symptoms of dependence or by standardized assessments, such as the Addiction Severity Index [ASI]), as compared with men. In particular, women tend to report greater psychological distress and mental health problems, particularly mood and anxiety disorders; more family-related needs, including issues related to parenting, child care, and child custody; more health-related problems; greater exposure to childhood and adult trauma and victimization and associated problems; and more problems related to lack of employment and vocational skills (Brady, Grice, Dustan, & Randall, 1993; Chatham, Hiller, Rowan-Szal, Joe, & Simpson, 1999; Stewart, Gossop, Marsden, Kidd, & Treacy, 2003; Wechsberg, Craddock, & Hubbard, 1998).
SEX DIFFERENCES IN TREATMENT UTILIZATION

Data on treatment admissions in the United States are reported into the Treatment Episodes Data System (TEDS) and provide valuable information on the extent to which women participate in treatment and differences in the characteristics of men and women who enter treatment. These data show that the overall proportion of men to women within the treatment system has remained fairly constant over the past 10 years (1995–2005) at 2:1, with women making up approximately one-third of all treatment admissions (OAS, 2006).

The sources of referral into treatment reveal the differential pathways through which women and men access substance abuse treatment. Nationally, a much higher proportion of men than women are referred into treatment through the criminal justice system (40 vs. 28%), whereas about twice as many women as men access treatment by referral from other community agencies (e.g., welfare, child welfare, health care providers; 15 vs. 6%; OAS, 2006). Other studies have also shown that women are more likely than men to enter treatment via the mental health and child welfare systems, whereas men are more likely to enter treatment through the criminal justice system (Schmidt & Weisner, 1995). Similarly, there are differences between men and women in their sources of payment for treatment, with a greater proportion of men reporting self-pay (26 vs. 18%) and a relatively greater proportion of women being dependent upon public insurance to pay for treatment (26 vs. 12%; OAS, 2006). These differences in source of referral and method of payment suggest that the pathways to treatment for men and women are strongly differentiated based on their economic and employment status, which may serve as either barriers or facilitators of treatment utilization. Moreover, the greater reliance of women upon public insurance to pay for treatment suggests that they are more vulnerable to changes in eligibility for, or reductions in, public insurance coverage for treatment.

In addition to the differential referral pathways into treatment, treatment participation is influenced by the severity of individuals’ substance use, their self-perception of the problems and adverse consequences related to their substance use, their access to resources, and the external forces that may support or inhibit treatment entry (Anglin, Hser, & Booth, 1987; Finney & Moos, 1995; Weisner & Schmidt, 1992). Some studies have shown that men and women define their substance use differently, in ways that either hinder or facilitate their treatment entry. In a study of 50 individuals with alcohol dependence in outpatient treatment, half of whom were women, the women frequently failed to identify their drinking as problematic and were adverse to being labeled “alcoholic” (Thom, 1986). They minimized the harmfulness of their drinking and emphasized that their drinking had not impaired their ability to fulfill their roles as wives and mothers. In fact, for many women, drinking with their spouses was integral to their marital relationships, and their spouses contributed to defining their drinking as nonproblematic. In contrast, nearly all of the men in the study who were living with a spouse or partner reported that they had been encouraged or supported by that person to enter treatment (Thom, 1987). Similarly, another study showed that a greater number of negative social consequences related to drinking—such as traffic accidents, arrests, and family-, legal-, and work-related problems—were associated with treatment entry among men but not among women (Weisner, 1993). However, employment and prior treatment history were related to treatment entry for both men and women.

These prior studies were confined to people with alcohol dependence; traditionally, women with alcohol dependence were relatively invisible, compared with men, as they risked being stigmatized as sexually deviant and social outcasts (Blume, 1986). Hence, their problems with alcohol may have remained relatively hidden and therefore triggered no social pressures to enter treatment. Moreover, traditional sex roles and marital relationships may have
inhibited treatment participation among women with alcohol dependence. There are similar findings with respect to women with drug use disorders. A study of couples in methadone treatment in the 1980s showed that patterns of drug use and decisions to enter treatment were often made jointly (Anglin, Kao, Hartlow, Peters, & Booth, 1987). More recent research has shown that a greater adherence to traditional sex roles among men in methadone treatment is positively related to psychological dominance and “couple drug involvement,” which in turn is associated with their perpetration of intimate partner violence (El-Bassel et al., 2004). Men in relationships where both partners used illicit drugs were more than twice as likely as others to perpetrate intimate partner violence and nearly four times as likely to perpetrate “severe” forms of violence (El-Bassel, Gilbert, Wu, Chang, & Fontdevila, 2007). Hence, the threat of intimate partner violence may limit women’s ability to enter treatment if drug use is integral to the couple’s relationship.

A study conducted with primary drug users demonstrates the contrary influences of marital relationships and social stigma on treatment participation for men and women. Using data from the national Drug Abuse Treatment Outcome Studies, Grella and Joshi (1999) examined the factors associated with having a history of substance abuse treatment among individuals who were sampled from residential, outpatient, hospital inpatient, and methadone maintenance programs. More severe drug use history and greater involvement in criminal behavior were related to prior treatment participation for both men and women. Yet there was a divergence between men and women in the other characteristics that were related to a history of treatment participation. Prior drug treatment among men was associated with higher levels of family opposition to their drug use and more support for their treatment participation. Treatment history among men was also associated with having been referred to treatment by their family, an employer, or the criminal justice system. In contrast, treatment history for women was associated with having been referred by a social worker, having a diagnosis of antisocial personality disorder, having engaged in sex work, or having initiated treatment on their own.

Taken together, this profile suggests that women’s treatment participation may be triggered by the greater “deviance” ascribed to their drug use and accompanying behaviors, such as sex work. Hence, women’s participation in treatment is inextricably bound up with the greater stigma that is associated with their substance use, vis-à-vis men’s. This distinction in social perception and consequences of drug use between men and women becomes even more salient when examining the factors that influence treatment participation among women who are pregnant or mothers of small children.

Given the greater centrality of their role as caregivers for children, the potential effects of treatment participation (or nonparticipation) on their relationships with their children may be a central influence on their decisions regarding whether or not to enter treatment. Several studies have shown that most women entering into substance abuse treatment are mothers of dependent children, and at least half have had contact with child welfare (Connors et al., 2004; Grella, Scott, Foss, Joshi, & Hser, 2003). However, less than half of mothers entering treatment are living with all of their children, and up to one-third have lost their parental rights to at least one child (Knight & Wallace, 2003; Schilling, Mares, & El-Bassel, 2004; Tracy & Martin, 2007). A study utilizing state administrative data showed that among mothers with a history of injection drug use, those who were residing with their children were significantly more likely to enter into methadone maintenance treatment (vs. other forms of treatment) than mothers who did not reside with their children (Lundgren, Schilling, Fitzgerald, Davis, & Amodeo, 2003). Mothers entering into methadone maintenance were also more likely to be employed and stably housed, compared with others, underscoring
the importance of access to resources in their ability to care for children as well as participate in treatment.

The influence of children upon parental substance use and mothers' motivation for, and participation in, treatment is complex (Collins, Grella, & Hser, 2003). Children (and pregnancy) may impede treatment participation if women fear it will jeopardize their retaining custody of their children (Haller, Miles, & Dawson, 2003). This factor was evident in a study in which involvement with child welfare was negatively associated with treatment motivation among mothers (Wilke, Kamata, & Cash, 2005). The authors suggested that the negative influence of child-welfare involvement on treatment motivation reflects the practical or emotional issues associated with having to leave children with other caregivers, as well as the fear of losing custody of their children. Similarly, another study found that among mothers who were opiate users, greater parenting responsibility, defined as number of children, was inversely related to number of treatment episodes, although this relationship was moderated by ethnicity and relationship status (McMahon, Winkel, Suchman, & Luthar, 2002). Further, many mothers claim to adopt strategies to limit the effects of their drug use on children, including restricting their use to certain time periods and hiding or disguising their use, thus discounting the harmful effects upon themselves and their children (Baker & Carson, 1999).

Logistical issues and access to resources also facilitate or impede treatment use among women. In a study of women with cocaine dependence and criminal involvement, both in and not in treatment, Saum, Hiller, Leigey, Inciardi, and Surratt (2007) showed that enabling factors, such as being legally employed, having health insurance, having custody of children, and knowing where to go to get treatment appeared to be the most influential predictors of treatment participation (as compared with predisposing and service need factors). In a study of individuals seeking publicly funded treatment in Washington State, women were found to spend longer on a wait list for treatment and to be less likely to eventually enter treatment, compared with men (Downey, Rosengren, & Donovan, 2003). The authors noted that pregnant women, who were given priority for admission to treatment, were excluded from the analyses; hence, women who were not pregnant may have had longer wait times stemming from the remaining limited capacity of inpatient and residential beds for women. They further surmised that women may postpone treatment entry due to lack of child care arrangements, thus prolonging their time on the wait list.

A qualitative study conducted in Australia examined barriers to treatment participation among 32 women who had sustained recovery for more than 1 year without participation in either treatment or self-help groups (Copeland, 1997). The sample identified the following barriers to treatment: social stigma and labeling, lack of awareness of the range of treatment options, concerns about child care, the economic and time costs of residential treatment, concerns about the confrontational approaches that were pervasive in traditional substance abuse treatment, and stereotyped perceptions of treatment services (e.g., the “religious” nature of 12-step groups). Moreover, several studies have shown that women with co-occurring mental disorders and/or who are homeless may be even more reluctant to enter into treatment because of their greater vulnerability, which is often related to histories of trauma and victimization (Padgett, Hawkins, Abrams, & Davis, 2006; Watkins, Shaner, & Sullivan, 1999).

Because of women's multiple vulnerabilities, some have argued that outreach and engagement strategies need to address the complex array of treatment needs that typically accompany substance use among women (Melchior, Huba, Brown, & Slaughter, 1999). Brown, Melchior, Panter, Slaughter, and Huba (2000) posit that women's readiness to make life changes must be assessed across four domains: domestic violence, HIV sexual risk behavior,
substance abuse, and mental health. Their resultant help-seeking behaviors reflect a hierarchy of readiness based on the immediacy, or time urgency, of their treatment issues across these domains. Moreover, previous studies have shown that experiences of socioeconomic disadvantage, exposure to community violence, criminal justice system interactions, and access to resources among women vary by ethnicity and influence perceptions of treatment needs and coping behaviors (Amaro et al., 2005, 2007).

PATHWAYS TO SUBSTANCE ABUSE TREATMENT THROUGH OTHER SERVICE DELIVERY SYSTEMS

Since the mid-1980s, major policy initiatives have influenced the pathways into substance abuse treatment for women through other service systems, either through mandated treatment, interventions to screen for substance abuse problems and refer clients to treatment, or cross-system collaborations. These policy changes and interventions are briefly reviewed.

Health Services

The primary health care sector is one of the predominant sources of treatment for individuals with alcohol or drug problems. Findings from the National Comorbidity Survey Replication showed that among individuals with a past-year SUD (either abuse of, or dependence on, alcohol and/or drugs), 26.2% received treatment from the mental health sector (including specialty substance abuse providers), and 18.1% received treatment from the general medical sector, which included treatment from physicians, nurses, and other health care professionals (Wang et al., 2005). Moreover, there was a dramatic increase in treatment for psychiatric disorders (including SUDs) in the general medical sector from 1990 to 2003, and women were more likely than men to be treated in this sector (Kessler et al., 2005).

Patient interactions with health care clinicians provide an important opportunity for screening and referral; hence, there is increased emphasis on developing screening tools and brief interventions for SUDs that can be implemented in primary care settings. In one study individuals screened for SUDs in a hospital emergency department were successfully referred to treatment and other services, including for women's health services such as breast cancer screenings and gynecological exams (Bernstein, Bernstein, & Levenson, 1997). Women who seek emergency medical treatment as a result of intimate partner violence, in which substance abuse is often involved, can be screened and referred to substance abuse treatment, as well as to other needed services (Lipsky, Caetano, Field, & Bazargan, 2005).

An optimal site for screening and referral of women for substance abuse interventions is within reproductive health services, such as obstetrics (Morse, Gerhson, & Hutchins, 1997). One study demonstrated the feasibility of using a brief screener to detect alcohol use among pregnant women in waiting rooms in eight obstetric clinics; 13% scored above the cutoff score indicating at-risk alcohol use (defined as binge drinking or more than one standard drink per week). At-risk use was predicted by smoking and earlier stage of pregnancy, indicating the feasibility of intervening early in pregnancies to reduce potential harm (Flynn, Marcus, Barry, & Blow, 2003).

Criminal Justice System

There has been an influx of women with substance abuse problems into the criminal justice system in the past 20 years, due to changes in sentencing and criminal justice policies that
have increased incarceration rates for drug users. Since 1995 the total number of female prisoners in the United States has grown by 53%, as compared to 32% for males (Harrison & Beck, 2005). Much of this increase can be accounted for by changes in sentencing for drug-related crimes, which have disproportionately affected women, particularly women of color (Freudenberg, 2002). As of 2002, 31.5% of female inmates nationally were incarcerated for drug-related crimes, compared with 20.7% of males. Thus, the criminal justice system is increasingly a conduit into substance abuse treatment for both men and women, including in prison-based settings and in community-correctional programs (Grella & Greenwell, 2004). As with women who enter into community-based treatment, women who enter into the criminal justice system typically have multiple service needs (Aleman, 2001). These include a greater likelihood of psychiatric disorders, particularly mood and anxiety disorders, compared with male offenders, as well as compared with women in the general population (Jordan, Schlenker, Fairbank, & Caddell, 1996; Pelissier & Jones, 2005; Teplin, Abram, & McClelland, 1996). Female offenders face substantial barriers to obtaining needed services, both in the community and in correctional settings. In one study female offenders were more likely to receive services for mental health needs in prisons than in the community preceding incarceration (Blitz, Wolff, & Paap, 2006). The lack of integrated treatment for female offenders with SUDs and co-occurring mental disorders is especially critical, given that these women are at higher risk for recidivism following their return to the community (Sacks, 2004).

Welfare

Legislation enacted in 1996, generally referred to as “welfare reform,” instituted Temporary Assistance for Needy Families (TANF). This state block grant program established federally mandated work requirements and a maximum lifetime 5-year limit to cash aid for clients. Along with these restrictions and requirements, the legislation instituted new expectations for local and state welfare systems, including the option to require screening for substance abuse and referral to treatment services as a condition of participation (Schmidt & McCarty, 2000).

Studies have shown that substance use is more prevalent among welfare recipients, compared with other women with dependent children who do not receive public assistance; however, the self-reported prevalence of illicit drug use is less pervasive than commonly assumed. Using national survey data, Pollack and colleagues showed that 22.3% of female TANF recipients reported illicit drug use, compared to 12.8% of women with dependent children who did not receive TANF (Pollack, Danziger, Jayakody, & Seefeldt, 2002). Among those who need treatment, however, rates of treatment are low. One study showed that fewer than half of TANF participants who were diagnosed with substance dependence at baseline participated in treatment over a 2-year follow-up period (Atkinson, Brown, Montoya, & Bell, 2004). Administrative barriers to the screening, assessment, and referral of drug-dependent recipients to treatment may impede their ability to access and utilize treatment services. Commonly noted barriers are the lack of resources and expertise among caseworkers to screen and assess for SUDs and the limited capacity of treatment systems to provide the comprehensive services needed by these recipients (Metsch & Pollack, 2005).

Although women with SUDs constitute a minority of TANF recipients, they face additional barriers to attaining economic self-sufficiency and are less likely to transition successfully to paid employment than recipients without SUDs (Jayakody, Danziger, & Pollack, 2000; Schmidt, Zakbiewicz, Jacobs, & Wiley, 2007). Given this population’s greater impairments and multiple problems, intensive case management services have been developed for
these women, which have increased their likelihood of entering and staying in treatment (Morgenstern et al., 2003). Moreover, welfare recipients with SUDs who stay in treatment longer and complete treatment are more likely to be working and not require welfare following treatment discharge, compared with those who do not complete treatment (Metsch, Pereyra, Miles, & McCoy, 2003).

**Child Welfare**

Greater awareness of the association between parental substance abuse and child abuse and neglect has fostered increased levels of interaction between these two systems in order to encourage coordination of services for child-welfare-involved parents with SUDs. Substance abuse treatment providers and child-welfare agencies are increasingly called upon to collaborate in providing services and making determinations of parental fitness and recommendations for child placement outcomes (Kerwin, 2005). However, historically these two service delivery systems have had differing orientations, goals, and organizational cultures, and these differences have created barriers to increasing coordination of services and case planning (Karoll & Poertner, 2002). Moreover, coordination of services across systems is further impeded when women fear that they may jeopardize custody of their children when they admit to substance abuse problems in child-welfare assessments or if they enter substance abuse treatment (Finkelstein, 1994; Jessup, Humphreys, Brindis, & Lee, 2003).

The development of dependency drug courts (or family drug courts), in which participation in treatment is mandated and supervised by a separate court or judge, has strengthened linkages between child-welfare and substance abuse treatment systems. Several recent evaluations have shown that women referred to substance abuse treatment through dependency drug court enter into treatment sooner and have higher rates of treatment completion, compared with those who are referred from child welfare but are not under court supervision (Boles, Young, Moore, & DiPirro-Beard, 2007; Green, Furrer, Worcel, Burris, & Finigan, 2007).

Although participation in substance abuse treatment may be made a condition of, or considered as a factor in, the determination of a woman’s parental rights, there is little understanding of whether or how participation in substance abuse treatment affects the outcomes of both parents and children within the child-welfare system. A study using state administrative data showed that women who were pregnant or who had custody of minor children were less likely than others to complete substance abuse treatment, although women who had children in foster care were more likely to do so (Scott-Lennox, Rose, Bohlig, & Lennox, 2000). Other studies have shown that mothers who are able to keep their children with them while in residential drug treatment (Hughes et al., 1995), or who retain custody of their infants while in intensive day treatment, have higher rates of treatment retention, particularly among those who are involved with child welfare (Chen et al., 2004) or who are mandated to treatment (Nishimoto & Roberts, 2001).

**INTERVENTIONS TO INCREASE TREATMENT PARTICIPATION: COERCION AND MOTIVATIONAL ENHANCEMENTS**

The effect of coerced participation in drug abuse treatment has typically been studied by addressing individuals who are involved with the criminal justice system, the majority of whom are men; the extent to which women are coerced into treatment has not been examined to the same extent. Over half of the sample of women participating in the national
multisite Women, Co-Occurring Disorders, and Violence Study reported that they had had at least one previous involuntary admission to alcohol, drug abuse, or psychiatric treatment, and 12% reported involuntary admissions to both psychiatric and substance use treatment (Clark et al., 2005). Yet there were no differences between women who had been mandated to treatment and those who had not, either in severity of substance use or psychiatric symptoms. The women who had been mandated to treatment, however, were more likely to have been arrested, suggesting that criminal behavior involvement, rather than clinical severity, is more likely to prompt such a mandate. Similarly, in a statewide treatment outcome study in California, women who were involved with the child-welfare system had lower levels of alcohol use severity and had no differences in psychiatric severity, but who were more likely to have been referred to treatment by the criminal justice system, compared with mothers with no child-welfare involvement (Grella, Hser, & Huang, 2006).

In a longitudinal follow-up study of a cohort that had initially been sampled at the point of treatment referral, external mandate was a much more powerful factor influencing treatment reentry among men over a 6-year follow-up period; there was a 12-fold greater likelihood of moving from using to treatment for men who were mandated to treatment, compared with women (Grella, Scott, Foss, & Dennis, 2008). Yet the same study showed that self-help participation was more strongly associated with moving from using to recovery for women. This finding is consistent with the findings from a longitudinal study of individuals who sought help for alcohol problems, showing that women were more likely than men to participate in self-help groups following initial treatment participation, and to have greater reductions in drinking concurrent with their self-help participation over an 8-year follow-up period (Timko, Finney, & Moos, 2005; Timko, Moos, Finney, & Connell, 2002). More recently, these findings have endured over a 16-year follow-up period (Moos, Moos, & Timko, 2006). Thus, whereas men may be influenced more to engage in treatment because of external mandates, women may be more willing to participate in self-help and other activities following treatment exposure.

Increasingly, the willingness of individuals with SUDs to enter and to stay in treatment has become a focus within treatment research. There has been considerable interest in recent years in “unpacking” the construct of treatment motivation and understanding its relationship to eventual treatment engagement and outcomes (Longshore & Teruya, 2006; Wild, Cunningham, & Ryan, 2006). Yet there has been remarkably little examination of sex differences in treatment motivation (Vasilaki, Hosier, & Cox, 2006), particularly in ways that can inform interventions that are designed to increase motivation and enhance treatment engagement and retention (Grella, 2008).

Motivational interventions use therapeutic strategies to increase the individuals’ awareness of their substance abuse problems and engage their commitment to behavior change. These interventions can build upon the issues that are central influences on women’s treatment participation, such as their identity, self-esteem, health concerns, and relationships with children, other family members, and friends. Moreover, motivational interventions can be tailored to address gender-related differences in differentiation and conception of self and coping styles in ways that increase women’s treatment participation (Cook, Epperson, & Gariti, 2005).

In most instances where motivational interventions have been developed specifically for women, they have aimed to promote changes in substance use among pregnant women with SUDs. Pregnancy is viewed as a “window of opportunity” in which women may be particularly receptive to substance abuse interventions that may be embedded with prenatal care (Handmaker & Wilbourne, 2001; Jessup & Brindis, 2005). In one example, a brief motivational intervention was used to address alcohol use among pregnant women in pri-
mary health care settings; an empathic, client-centered motivational interview focused on the health of participants' unborn babies (Handmaker, Miller, & Manicke, 1999). Participants who had the highest blood alcohol levels during early pregnancy had greater reductions at a 2-month follow-up if they received the motivational interview, compared with participants who received an informational pamphlet. Similarly, in another experimental study, pregnant women with the highest levels of consumption had significantly greater reductions in alcohol use after a single-session brief intervention, compared with a usual care group; furthermore, the effect was enhanced when the woman’s partner participated in the session (Chang et al., 2005). Yet in another study, non-treatment-seeking pregnant women were recruited from prenatal clinics and assigned to two conditions. Both groups received motivational interviewing combined with behavioral incentives, and one group received case management services in addition (Jones, Svikis, Rosado, Tuten, & Kulstad, 2004). Although the addition of case management increased access to needed services, there was no effect of the combined motivational interviewing and behavioral incentives on number of counseling visits or completion (defined as four visits). The authors concluded that the combined intervention may have lacked sufficient intensity to engage these participants in treatment.

**FUTURE DIRECTIONS**

Building upon this large and growing body of research, future research should continue to track the patterns of treatment and service utilization among women with substance use problems. It will be especially important to understand the factors that facilitate or impede access to treatment for women of differing characteristics, including parental status, and the interventions that increase screening, referral, and treatment entry across different settings. A full understanding of women’s treatment utilization requires the integration of epidemiological data on substance use prevalence and need for treatment, health services research on the pathways into treatment and the costs of different service configurations, clinical research on the characteristics of women who enter treatment and their treatment outcomes, and longitudinal studies that track patterns of substance use, treatment utilization, and recovery over time. Such a research base will provide the foundation necessary for fully informed policy initiatives that prioritize substance abuse treatment for women who need it.

**KEY POINTS**

- The historical legacy of greater stigma associated with alcohol and drug use among women, and the relative invisibility of women’s treatment needs, compared with men’s (Kandel, 1996), have influenced the pathways into treatment and treatment options available to women.
- As women’s alcohol and drug use has neared parity with that of men, particularly among users of stimulants, or, in the case of prescription drug abuse, exceeded that of men, women with substance abuse problems have increasingly entered into treatment, as well as the criminal justice system.
- Heightened public awareness of the potential consequences of maternal substance use on children has focused attention on alcohol and drug use among pregnant and parenting women. The spotlight on mothers, however, has often promoted greater stigma and increased barriers to treatment participation among women who fear incarceration or loss of custody of their children.
- At the same time, policy initiatives focusing on maternal substance abuse have increased fund-
ing and treatment capacity dedicated to pregnant and parenting women. (Grella & Greenwell, 2004; Schmidt & Weisner, 1995). An unintended consequence of this policy focus on pregnant and parenting women is that, other women, such as older women, may have less access to treatment. However, there has been little investigation into the treatment needs of this population thus far. (Hamilton & Grella, 2008).

- Women with SUDs frequently come into contact with other service systems, including mental health, health services, welfare, child welfare, and criminal justice. Various interventions to screen women for SUDs, to refer them to appropriate treatment, and to coordinate delivery of services have been implemented across these systems.

- Future research can build upon this existing body of research on women's treatment utilization by continuing to examine the relationships of treatment need to access and utilization, longitudinal patterns of service use and outcomes, and organizational and system-level factors that improve the delivery of treatment services to women.

REFERENCES

Asterisks denote recommended readings.


Schmidt, L., & Weisner, C. (1993). The emergence of problem-drinking women as a special population


Chapter 25. Effects of Work-Family Policies on Parenthood and Wellbeing
Caitlyn Collins and Jennifer Glass

<1> Introduction

Becoming a parent makes adults happier and healthier—or does it? While popular cultural beliefs bolster this claim, a wealth of studies beginning in the 1970s find that overall mothers and fathers typically report lower levels of wellbeing than childless adults across industrialized countries. This finding emerged in the context of substantial economic and social changes in work and family life over the past four decades: an increase in women’s labor force participation, a decline in men’s earnings, and a rise in dual-earner and single parent families (Kohler et al., 2006; McLanahan and Adams, 1989). The finding that nonparents report greater happiness than parents presents a puzzle for researchers because decades of scholarship suggest the opposite: since Durkheim’s (1897) classic study of suicide, research has found repeatedly that social roles and relationships such as parenting have positive benefits for adults’ mental health (House et al., 1988; Mirowsky and Ross, 2003). This anomalous finding is a relevant and timely policy issue given its possible association with high rates of childlessness, stress, and work-family conflict for adults, and poorer outcomes for children (Aassve et al., 2005; Gornick and Meyers, 2003; Mather, 2012).

Some scholars have employed theories of mental health and stress to suggest that the emotional and financial costs of contemporary parenthood outweigh the emotional rewards of having children (Balbo et al., 2012; Begall and Mills, 2011; Evenson and Simon, 2005; Liefbroer, 2005; McLanahan and Adams, 1989; Nomaguchi and Milkie, 2003; Woo and Raley, 2005). Yet the experience of contemporary parenthood varies widely from country to country depending on the sorts of resources and social supports offered to parents (Esping-Anderson, 1990; Kahneman et al., 2010; Savolainen et al., 2001). For example, Scandinavian countries offer extensive supports and socialize the cost of childrearing across society, while some Mediterranean countries and the United States offer minimal supports, compelling parents to find private solutions to meet their families’ needs (Gornick and Meyers, 2003 and Chapter 5 by Eydal et al. and Chapter 10 by
Thevenon, in this volume). Parenthood can be particularly taxing and stressful in countries that offer little support to parents (Kahneman et al., 2010), and scholars have long suggested that increased institutional support such as readily available childcare would greatly improve mothers’ and fathers’ psychological wellbeing (Bird, 1997).

Although recent scholarship has begun examining cross-national variation in the relationship between parenthood and wellbeing (Kahneman et al., 2010; Ono and Lee, 2013; Savolainen et al., 2001), we still know little about why the parenthood gap in wellbeing is larger in some countries than in others. We suggest that the parenthood gap in wellbeing is associated with state-provided public policies supporting families, and that this is the most promising avenue for investigating country-level differences in parental status differences in happiness. In this chapter, we first review the literature on parental wellbeing in Organization for Economic Cooperation and Development (OECD) countries. Second, we discuss the findings of studies that have considered whether, and to what extent, work-family policies lessen the time costs, financial costs, and psychosocial stress associated with parenting. We provide several empirical examples from our research. Third, we conclude by suggesting avenues for future research that help move the work-family policy agenda forward by understanding the relationship between macro-level contexts and micro-level emotional processes for parents in OECD countries.

<1> Parental Wellbeing in Advanced Industrialized Countries

Across developed countries, considerable research has documented a significant negative association between parenthood and emotional wellbeing (Gilbert, 2007; Hansen, 2012; Kahneman et al., 2004; McLanahan and Adams, 1989; Nomaguchi and Milkie, 2003; Simon, 2008; Stanca, 2012; Umberson et al., 2010). This association has been found in all household types, for both mothers and fathers, and across a number of dimensions of emotional wellbeing—including depression, anxiety, generalized distress, life satisfaction, and the frequency of everyday positive and negative emotions like happiness and anger (Ross and Van Willigen, 1996; see Hansen, 2012, Umberson et al., 2010, and Nelson et al., 2013 for recent reviews). Yet the personal and household characteristics of parents also matter (Aassve et al., 2012; Umberson et al., 2010; Woo and Raley, 2005). Although married and cohabiting parents report less stress than single parents (Aassve et al., 2012; McLanahan, 1983; Simon, 1998; Woo and Raley, 2005), research shows consistently that across marital statuses, parents residing with minor children—whose time, energy, and financial demands are highest—show lower levels of wellbeing than adults who do not reside with children (Evenson and Simon, 2005).
What about older parents whose children have left home? Common reasoning suggests that empty-nest parents might be happier than childless adults since the emotional benefits of parenthood are thought to be greatest when children have grown up and become independent. However, research in the United States does not support this claim: empty-nest parents do not enjoy greater emotional wellbeing than childless adults (Bures et al., 2009; Evenson and Simon, 2005; Koropeckyj-Cox, 2002; Milkie et al., 2008; Pudrovska, 2009). In fact, Evenson and Simon (2005) found that there is no type of parent (including custodial, non-custodial, and step-parents of both minor and adult children) reporting significantly better mental health than nonparents in the United States.

Why is parenthood so often associated with lower levels of emotional wellbeing? Parenthood, like other major adult social roles, provides individuals with an important sense of meaning and identity, personal gratification, and social connections to others, which improve mental health (Nomaguchi and Milkie, 2003; Umberson and Gove, 1989; Umberson et al., 2013). But the emotional rewards of parenthood could be overshadowed by the stress of raising children. Theories operationalizing how stress affects mental health (Pearlin, 1989) have the greatest potential to explain the parenthood gap in emotional wellbeing. Because children increase adults’ exposure to a number of stressors, parenthood increases symptoms of anxiety, depression, distress, and anger while decreasing positive emotions such as happiness.

Fawcett (1988) outlined three major types of burdens parents confront that undermine their emotional wellbeing: time costs, financial costs, and psychosocial stress. First, parents who live with minor children experience substantial demands on their time and energy coupled with sleep deprivation (Avison et al., 2007; Nelson et al., 2013) and work-family conflict (Begall and Mills, 2011; Nomaguchi et al., 2005). The increased amount of time spent doing housework and caring for children decreases leisure time and time with one’s spouse for married parents, which may decrease marital satisfaction (Claxton and Perry-Jenkins, 2008). In dual-earner households in which both parents work full time, parents experience high levels of time pressure, particularly in the U.S. (Mattingly and Sayer, 2006; Milkie et al., 2009; Nomaguchi, 2009).

Second, raising minor children at home is associated with increased financial strain (McCrate, 2005; Nelson et al., 2013; Warren and Tyagi, 2004). The added expense of having children can lead to economic hardship, which has negative repercussions for parents’ psychological wellbeing (Mirowsky and Ross, 2003). For example, parents of young children in many countries have difficulties obtaining high quality, affordable childcare (Bird, 1997; Kravdal, 1996; Ross and Mirowsky, 1988), while some parents of older children face the economic burden of financing children’s higher education and independent living at the same time they face their own retirement (Fingerman et al., 2012; Furstenberg et al., 2004; Warren and
Low-income parents are exposed to additional sources of financial stress, such as the stress of living in unsafe neighborhoods with under-resourced schools, unreliable and inadequate daycare and healthcare for their children, and insufficient food (Edin and Kefalas, 2005; Heymann, 2000; Nomaguchi et al., 2005; Ross and Mirowsky, 1988).

Third, parents experience psychosocial stress from the daily ins and outs of parenting. Prolonged psychosocial stress is problematic because it can lead to poor performance, chronic fatigue, disinterest, dejection, lack of work motivation, memory and sleep disturbance, numbness, muscle pain, recurrent infections, and to chronic conditions like depression, diabetes, and cardiovascular disease (Danielsson et al., 2012). Some research shows that women tend to experience greater stress and stronger negative shocks to wellbeing in the transition to parenthood (Nomaguchi and Milkie, 2003; Simon, 1992). Given that mothers still complete the majority of childrearing and housework in all industrialized countries, even when employed full-time (Pettit and Hook, 2009), studies have explored whether wellbeing differs for mothers versus fathers in dual-earner households. The majority of studies show few differences (Evenson and Simon, 2005; Margolis and Myrskyla, 2011). While mothers report more work-family conflict (Gornick and Meyers, 2003) and less leisure time (Mattingly and Bianchi, 2003) than fathers do, fathers report more distress about the financial strains of parenting and the tension between breadwinning and spending time with children (Aumann et al., 2011; Simon, 1998).

Some aspects of parenting, like multitasking, are not experienced equally by mothers and fathers. Mothers spend on average 10 hours more on multitasking per week than fathers, primarily on housework and childcare. These multitasking activities for mothers at home and in public lead to increased stress, negative emotions, work-family conflict, and psychological distress (Offer and Schneider, 2011). At the same time, fathers’ multitasking at home, which involves less housework and childcare, is not considered a negative experience. A survey of 909 working women in Texas found that childcare ranked 16th out of 19 possible pleasurable activities (Kahneman et al., 2004), behind watching television, preparing food, napping, housework, shopping, and talking on the phone. These stressors, unsurprisingly, are more acute for single parents than for married and cohabiting parents (Aassve et al., 2012; Avison et al., 2007; McLanahan, 1983; Meadows et al., 2008; Simon, 1998), helping explain why single parents report the lowest levels of emotional wellbeing of all parents. Simon and Caputo (n.d.) find that although adults experience greater emotional wellbeing and less emotional distress at different stages of parenthood than others, overall, they find no overall advantages of parenthood for adults’ health and happiness.

The majority of scholarship to date has focused on the proximate sources of stressors that mediate the association between parenthood and emotional wellbeing, such as sleep deprivation
and financial strain. Nelson and colleagues (2013) present a parent wellbeing model that maps the psychological mechanisms that mediate this relationship, outlining both why and how parents experience more or less wellbeing than nonparents. Their model identifies both positive mediating mechanisms (purpose/meaning in life; human needs; positive emotions; social roles) and negative ones (negative emotions; financial strain; sleep disturbance; strained partner relationships).

In this chapter, we extend this model by zooming out to highlight the distal sources of stress – like childcare and parental leave – that are rooted in the larger policy context in which adults raise children. Prior research suggests that these distal factors help mediate the relationship between parenthood and wellbeing. For example, Bird (1997) finds that children in the U.S. do not increase the psychological distress of parents in and of themselves: rather, children are associated with greater social and economic burdens, which increases parents’ levels of distress. She argues that if parents were relieved of the economic hardship associated with having children by providing more material support to parents, they would not display significantly higher distress levels than nonparents.

Given that parenting is found to have both positive and negative effects on wellbeing, “it should be possible to decrease the psychological costs of parenting by reducing the burdens or constraints that overshadow the more rewarding aspects of raising children” (Bird, 1997: 810). Research supports this claim: publically provided benefits such as a statutory right to reduced working hours when children are young and cash allowances to parents are shown to reduce some of the stressors such as very long work hours and financial distress associated with raising children to adulthood in industrialized countries (Gornick and Meyers, 2003; Heymann et al., 2007). We therefore theorize parenting as a stressor buffered by institutional support (Glass et al., 2016), and argue that attention to the larger sociopolitical context—particularly different work-family policies—provides further insight into why the emotional effects of parenthood vary from country to country. We present a modified and expanded version of Nelson et al.’s model of proximate sources of stress (see Figure 25.1) to highlight the distal policy supports that might improve parents’ wellbeing. We are interested specifically in how work-family policies might affect the psychological mechanisms that mediate the relationship between parenthood and wellbeing. For example, how might paid parental leave and subsidized childcare impact the mediating relationship of financial strain and work-family conflict on the wellbeing of parents versus nonparents?

<INSERT FIGURE 25.1 HERE>
The welfare state regimes of advanced industrialized countries offer policy packages that vary widely in their levels of social support and resources to parents (Esping-Anderson, 1990; Kahneman et al., 2010; Savolainen et al., 2001). In the social-democratic regimes of Scandinavia, policies such as extensive childcare and workplace accommodations seek to reduce the direct financial costs and opportunity costs of parenthood, equalize the household division of labor, and ameliorate work-family conflict and overload (Gornick and Meyers, 2003, see also Chapter 15 by Eydal et al., in this volume). Prior studies have identified these costs as the main sources of the parenthood deficit in emotional wellbeing across industrialized countries (Pollmann-Schult, 2014; Stanca, 2012). Within Europe, parents living in social-democratic countries tend to be happier than those living in countries with liberal or conservative welfare regimes (Aassve et al., 2012). In conservative welfare state regimes such as Germany, policies have long tended to support a male breadwinner/female homemaker model of family life. On the other end of the spectrum, few policy supports are available for families with children in liberal welfare state regimes such as the U.S. and some Mediterranean countries. In these regimes, parents contend with the demands of childrearing with their own resources and social networks (Glass, 2000; Gornick and Meyers, 2003; Simon, 2008, see also Chapter 16 by Naldini et al., Chapter 13 by Pfau-Effinger and Chapter 14 by Woods in this volume).

In terms of wellbeing, women in the social-democratic welfare states seem to suffer the least as a result of childbearing, whereas women in conservative and Mediterranean states suffer significantly more. For liberal welfare regimes the results are more mixed, and depend on the definition of wellbeing used (Aassve et al., 2005). The U.S. in particular is exceptional in its lack of policy provisions to offset the costs of childrearing and work-family conflict (Gornick and Meyers, 2003). Only one federal policy—the 1993 Family and Medical Leave Act—is designed to help people meet the dual demands of work and family. In a study of 22 OECD countries, the U.S. had the largest subjective wellbeing penalty for parenthood (Glass et al., 2016).

Although broad welfare state typologies are useful to help understand the relationship between parenthood and wellbeing cross-nationally, we suggest that an examination of specific public policies intended to reduce parental stressors better illuminates which policy contexts alleviate the gap in wellbeing between parents and non-parents in OECD countries. Next, we review findings on which policies help reduce the time costs, financial costs, and psychosocial stress associated with parenthood.
Vacation and sick day policies help parents preserve time away from work to care for and spend time with children (Gornick and Meyers, 2003), and have been shown to dramatically improve wellbeing among parents (Heymann et al., 2007; Hyde et al., 1995). Using data from 22 OECD countries, Glass and colleagues (2016) report a strong policy effect of vacation and sick days on parental happiness. Paid parental or maternity leave also enable parents of newborns to leave work temporarily to care for an infant. These policies together reversed the negative impact of parenthood on happiness (Glass et al., 2016).

Work schedule flexibility policies allow workers to organize their time in a way that makes the combination of employment and parenting feel more compatible (see Chapter 11 by den Dulk, this volume). Three types of work schedule discretion tend to be available for employees: the ability to change the starting and ending times of the workday, the ability to take time off during the work day to attend to personal matters, and the ability to refuse overtime work (Golden et al., 2013). In countries like Denmark and Sweden, strong work-time policies allow parents to temporarily reduce their working hours or set maximum weekly work hours, and in the UK, “right-to-ask” laws support flexible work schedules (Hegewisch and Gornick, 2008). Golden and colleagues (2013) find a significant association between schedule flexibility and workers’ reported happiness—particularly the opportunity to take time off during the work day, and to a lesser extent, to vary daily starting and quitting times. Lower stress levels are also reported among employees who control their work hours and schedules (Grzywacz et al., 2008). Others find that both parents and nonparents benefit from living in countries with a larger percentage of workplaces offering flexible schedules, with nonparents benefiting slightly more (Glass et al., 2016). They also find that work flexibility does not increase happiness more for parents than nonparents.

<2> Policies Ameliorating the Financial Costs Associated with Parenthood

A large body of work suggests that parental wellbeing is strongly contingent on the financial costs borne by parents (Bird, 1997; Pollman-Schult, 2014). A comparative study of 94 countries finds that children’s negative impact on parents’ wellbeing is explained primarily by a large negative effect on household finances (Stanca, 2012). Although when approached globally, resident children are associated with decreased financial satisfaction for parents (Angeles, 2010), this is not the case in Nordic countries that have extensive work-family policy provisions (Hansen et al., 2009; Savolainen et al., 2001).
The cost and availability of childcare in particular has a strong influence on parents’ wellbeing. Glass and colleagues (2016) find that policies lowering average childcare costs shows the greatest potential to increase parental happiness. This is supported by Misra et al. (2007), who report that generous, high-quality childcare reduces financial strain and lowers poverty rates, particularly for single mother-headed families. In-depth interviews with middle-income working mothers in Sweden and the U.S. further support these findings: Swedish mothers did not report worrying about childcare costs, while American mothers reported enormous stress related to the high cost of daycare. U.S. mothers spent significant time researching and securing care they could afford, and some were forced to quit work when they couldn’t find affordable care (Collins, 2016). Without universal childcare like Sweden has for children over age one, no one daycare solution was reliable for American families—all were temporary arrangements that could shift unexpectedly. Daycare centers closed, babysitters started different jobs or moved, and relatives who helped out fell ill themselves. American mothers who had the most financial resources available to dedicate to childcare tended to be the happiest with the solutions they found. A wealth of research demonstrates that low-cost childcare improves parental wellbeing both by enhancing incomes and reducing stress (Ross and Mirowsky, 1988; Savolainen et al., 2001; Stanca, 2012).

Pollman-Schult (2014) reported that the financial costs of children negatively impact men’s life satisfaction in particular, especially when the mother has left the labor force. This echoes the finding that the perceived financial costs associated with parenthood have a strong effect on men’s entry into parenthood—negatively impacting countries’ fertility rates (Liefbroer, 2005). Together, these findings suggest that both men’s interest in starting a family and their life satisfaction after having children are deeply impacted by the financial aspects of parenthood (Pollman-Schult, 2014). More broadly, child benefits—which vary drastically from country to country—play an important role in ameliorating poverty for families (see Bradshaw, this volume).

In countries that offer little policy support to parents, parenthood is particularly emotionally taxing and stressful (Kahneman et al., 2010). Insofar as work-family policies (i.e. parental leave, paid vacation and sick days, flexible work schedules, and subsidized public childcare) give parents time to decompress, complete errands and household tasks, care for loved ones, and spend time both with and away from children, all of these policies help reduce parents’ stress and worry (Collins, 2016; Glass et al., 2016; Gornick and Meyers, 2003). Interestingly, Glass and colleagues
find that child allowances (cash benefits paid directly to families) show little impact on parental wellbeing, perhaps because the average cash benefit in most countries remains fairly low given the costs of raising children.

On the other hand, flexible work scheduling positively influences subjective wellbeing measures like work-life conflict, work stress, and fatigue (Golden et al., 2013). Paid parental leave is positively associated with parental happiness (Glass et al., 2016). Research also suggests that the availability and length of maternity leave is related to women’s mental health: women who take short leaves and have poor-quality marriages are most vulnerable to depression (Hyde et al., 1995). These paid childbearing leaves influence parental happiness likely because they strengthen later parent-child attachment and minimize the long-term employment costs of parenthood (Glass et al., 2016; Misra et al., 2011).

Childcare quality is also important for mothers’ mental health and wellbeing. Qualitative research suggests that the availability of high-quality, affordable childcare substantially increases working mothers’ sense of work-life balance and satisfaction in Sweden and Germany, while the absence of reliable childcare in the United States and Italy is a monumental source of stress for mothers who work outside the home (Collins, 2016). Gordon and colleagues (2011) report that mothers who choose their children’s daycare facility based on perceptions of quality report lower depressive symptoms than those who choose for practical reasons such as cost, hours, and location.

Overall, these findings suggest that policy differences between countries may help shape the balance of costs and rewards for parenthood (Glass et al., 2016; Hansen, 2012; Margolis and Myrskyla, 2011, 2015). Parents living in countries with the most generous work-family policies and greatest gender equality seem to derive the greatest emotional benefits from having children (Hansen, 2012). In fact, the inverse relationship between parenthood and happiness is completely eliminated in nations with the strongest policy packages—which with paid parental leave, paid sick and vacation leave, and work schedule flexibility (Glass et al., 2016). Because fertility tends to be low in countries where the costs of parenthood are particularly high (Hilgeman and Butts, 2009), work-family policies may be a mechanism by which governments can increase fertility by removing the disincentives to parenting, equalizing the cost of childrearing, and easing the combination of employment and parenthood (McDonald, 1997). Further, given the well-established relationship between parental and child wellbeing (Cummings et al., 2005), it is likely that work-family policies improve not only the wellbeing of parents, but also their children (Glass et al., 2016). Policies like parental leave influence the stress of caregiving and couple’s relationship stability, household division of labor, and children’s relations to their parents (Almqvist and Duvander, 2014). Even more significantly, welfare state policies seem to improve
the levels of happiness for countries’ general populations, both parents and nonparents alike (Flavin et al., 2014; Glass et al., 2016).

Glass and colleagues (2016) constructed a comprehensive policy index (CPI) for 22 OECD countries that combines work flexibility, paid vacation/sick leave, and paid leave available to mothers and tested whether the nations that score higher on this index have a smaller gap in parental happiness. They find that the countries that offer the strongest family policies (i.e. those that score highly on the CPI) exhibit a net positive effect of parenthood on happiness, while those that offer the weakest policies (i.e. those that score lower on the CPI) show a deficit in parental happiness compared to nonparents.

<INSERT FIGURE 25.2 HERE>

The extent of gender-specific responses to work-family policies has not been settled, but some studies suggest this possibility. The legacy of “breadwinner-housewife” gender ideologies can be seen in men’s responses to the financial demands of parenthood even as couples rely more heavily on women’s earnings to meet those demands (Pedulla and Thébaud, 2015). Employed women, by contrast, seem more affected by the time pressures and multitasking associated with parenting children, which disproportionately fall to mothers (Offer and Schneider, 2011). This suggests that men might be more affected by income supports for parenting, while women may respond more to time and scheduling interventions.

<1> Conclusions

We conclude by discussing the avenues that future research should explore to improve our understanding of how public policies impact parental wellbeing. The heuristic model by Nelson and colleagues (2013) includes only emotional/psychological wellbeing as an outcome, but objective measures such as financial debt, disease burden, and cortisol (stress hormone) levels could also productively be used to measure the micro-level impacts of social policies. Future research should measure these objective aspects of wellbeing and uncover their role in producing both low second birth rates and low subjective wellbeing among parents. Better integration of health data with time-use studies across the EU and English-speaking countries would also be helpful to see how policies change the daily patterning of activities and interactions in families of different household types. Scholars should especially focus on households disadvantaged by income, ethnicity, or family structure (number of adults in the household) since these are the households most impacted by public policy or the lack thereof. We also need further research that
investigates the role of partnership status (single, cohabiting, married), race and immigration status, age at parenthood and age of children, and social class on parents’ response to social policies.

While policy makers seem to excessively focus on child wellbeing, the problem of low fertility ultimately stems from the effects of policies on parents because, although we want children to benefit from policies, policies affect the context in which adults’ make decisions about having children. We are missing the larger picture: why do parents decide (not) to have children, especially a second child after the first has arrived? Is the policy framework enough to still make parenthood desirable or is it just ameliorating the worst aspects of it? A research paradigm that focuses on the policy mix producing the highest level of Gross Domestic Happiness in families might ultimately resolve the problem of below replacement fertility faster than a focus on either cash transfers or fertility incentives. Currently, the costs of parenthood look so enormous in comparison to its mainly psychosocial benefits that many individuals simply opt out of reproduction. When the time and fiscal resource drain of raising children induces a free-rider problem in the reproduction of the next generation of workers and citizens, policymakers must attend to the fairness of the cost distribution among parents, employers, and the state.

But this raises an important conundrum: that of how to increase policy support for parenting without harming those at other life stages. Some scholars (Ono and Lee, 2013) have suggested that social-democratic welfare states merely redistribute happiness among policy-targeted demographic groups in these countries. This suggests that the pro-family ideology of the social-democratic welfare states protects families from social risk and improves their well-being at the cost of single persons. However, Figure 25.2 suggests that the same policies that improve parental happiness also improve the happiness of nonparents, although to a lesser extent. Future research must determine the presence and extent of any redistribution of happiness between parents and nonparents, while simultaneously asking the thorny question of whether such redistribution is nevertheless in the best interests of children, families, and the larger social good.

References


between free time and feeling rushed’, *Journal of Marriage and Family*, 68 (1), 205–221.


Figure 15.1. Distal Supports and Proximate Stressors that Mediate the Relationship Between Parenthood and Wellbeing.
Figure 15.2. Estimated Happiness (1-10 scale) for Parents and Non-Parents with Comprehensive Policy Index (CPI), Source: ESS and ISSP, 2004-2007¹ (from Glass, J., R. Simon, and M. Andersson (2016), ‘Parenthood and happiness: Effects of work-family reconciliation policies in 22 OECD countries’, *American Journal of Sociology*).

¹ All control variables at the individual and country level are held at their sample mean
Perceived Need for Mental Health Care: The Intersection of Race, Ethnicity, Gender, and Socioeconomic Status

Alice P. Villatoro1, Vickie M. Mays2, Ninez A. Ponce2, and Carol S. Aneshensel2

Abstract
Racial/ethnic minority populations underutilize mental health services, even relative to psychiatric disorder, and differences in perceived need may contribute to these disparities. Using the Collaborative Psychiatric Epidemiology Surveys, we assessed how the intersections of race/ethnicity, gender, and socioeconomic status affect perceived need. We analyzed a nationally representative sample of U.S. adults (18 years or older; N = 14,906), including non-Latino whites, Asian Americans, Latinos, African Americans, and Afro-Caribbeans. Logistic regressions were estimated for the total sample, a clinical need subsample (meets lifetime diagnostic criteria for a psychiatric disorder), and a no disorder subsample. Perceived need varies by gender and nativity, but these patterns are conditional on race/ethnicity. Men are less likely than women to have a perceived need, but only among non-Latino whites and African Americans. Foreign-born immigrants have lower perceived need than U.S.-born persons, but only among Asian Americans. Intersectional approaches to understanding perceived need may help uncover social processes that lead to disparities in mental health care.

Keywords
perceived need, mental health care, race/ethnicity, gender, socioeconomic status

INTRODUCTION
Consistent variation in mental health services utilization across social strata can be indicative of mental health care disparities when these differences are due to access barriers and persist net of need for services. The limited use of mental health services among racial/ethnic minority groups relative to non-Latino whites is of critical concern because these disparities remain even after adjusting for clinical need and other dimensions of social status (e.g., Abe-Kim et al. 2007; Alegrı́a et al. 2002, 2008; Jackson et al. 2007; Neighbors et al. 2007). Existing research on mental health care disparities examines the role of policies, provider practices, and differential access, often overlooking attitudes and perceptions about mental illness and its treatment. The presence of a psychiatric disorder, a commonly used but imperfect indicator of need for services (Druss et al. 2007; Mechanic 2003; Tuithof et al. 2016), is a major influence on help-seeking behaviors (Alegrı́a et al. 2007; Jang et al. 2015; Wang et al. 2005). Nevertheless, many persons with psychiatric illness may not be aware of the need for services or may not understand the nature of their illness. Understanding the presence of need and the factors that influence perceived need may help to identify social processes that contribute to disparities in mental health care.

1Columbia University, New York, NY, USA
2University of California, Los Angeles, CA, USA

Corresponding Author:
Alice P. Villatoro, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 W. 168th St., 7th Floor, Room 720C, New York, NY 10032, USA.
Email: apvilla@gmail.com
disorders, including serious mental illness, do not self-label as having a disorder, much less one that requires treatment (Kessler et al. 2001; Mojtabai et al. 2011). Perceiving a need for mental health care, an element of the self-labeling process, is a pivotal precursor to voluntary help-seeking (Mojtabai et al. 2011; Mojtabai, Olfson, and Mechanic 2002; Narendorf and Palmer 2016) and is thought to contribute to racial/ethnic disparities in mental health care utilization (Nadeem, Lange, and Miranda 2009), but little is known about the social distribution of perceived need, especially between and within racial/ethnic groups. The current study examines perceived need as a type of self-labeling and the extent to which it varies across lines of social demarcation, focusing on the intersections of race/ethnicity with two other status characteristics with strong links to both mental health and treatment—gender and socioeconomic status (SES).

**Perceived Need for Mental Health Care**

We conceptualize perceived need for mental health care as a person’s awareness that something is wrong, that what is wrong relates to his or her mental health, and that professional assistance is necessary to overcome it. This perception figures prominently in sociological theories about help-seeking. The widely used Behavioral Model of Health Services Use, for example, posits three main influences on help-seeking behavior: predisposing characteristics, enabling factors, and need, which includes actual (biological/clinical) and perceived need (Andersen and Davidson 2007). Correspondingly, perceived need is a key element of mental illness career models (Aneshensel 2013; Karp 1996), self-labeling theory (Thoits 1985, 2005), and the network-episode model, which describes the impact of social support systems and treatment systems on how individuals view their mental health and whether they see themselves as requiring help (Pescosolido, Boyer, and Medina 2013).

Perceived need is conceptualized here as an element of self-labeling theory (Thoits 1985, 2005), which seeks to explain how individuals perceive and label their mental health problems. This theory posits that individuals experiencing significant emotional distress are likely to recognize that their psychological problems require professional treatment; this self-labeling process then motivates individuals to voluntarily seek help or treatment, although not all do so, and in some cases “voluntary” treatment may not be entirely voluntary (Thoits 2005, 2011). As summarized by Moses (2009), Thoits (1985) asserts that three criteria are necessary for self-labeling to occur: (1) The individual is assumed to be a well-socialized actor who shares the cultural perspectives of others and understands normative expectations; (2) norms about acceptable behaviors are clear, known, and can be applied to oneself or others; and (3) the individual is motivated to conform to social expectations. As such, self-labeling requires the individual to have “knowledge of and acceptance of basic psychological dynamics that underlie the etiology of mental disorder” and sufficient material resources to recognize and seek care (Thoits 2005:105). Therefore, these criteria suggest that variation in cultural orientations, adherence to dominant social norms, and location in the social system may generate variation in self-labeling across racial/ethnic groups, especially in relation to non-Latino whites.

The awareness that something is wrong and requires treatment often begins with the attribution of inchoate feelings of distress to proximal events and circumstances (e.g., job loss), which subsequently becomes identified as an internal dysfunction when the situation changes but the distress persists (Karp 1996). Thoits (2011:4) cites the person’s self-awareness of emotional reactions to stressful situations as “inappropriately intense, recurrent, or prolonged.” In sociological terms, this realization signifies onset of illness because it is the perception of the condition as distinct from any underlying biological malfunction (Aneshensel 2013).

However, Thoits (2005) notes that differential self-labeling across social status groups may exist because socially and economically advantaged groups (e.g., non-Latino whites, high SES) will find it “easier” to self-label than others as a result of better mental health literacy and possession of sufficient resources to seek care. For instance, self-labeling is more prevalent among high SES than low SES adults, resulting in relatively high rates of voluntary service use (Thoits 2005) and relatively low probabilities of deflecting a mental illness identity or label (Thoits 2016). Moses (2009) demonstrates that self-labeling is a difficult process for youth, even for those who are in treatment: Only some used psychiatric terms to label their problems, while others applied different labels or none at all. Among these youthful patients, self-labeling was associated with being
white relative to being a member of a racial/ethnic minority group; however, generalizability is limited because the study used a convenience sample of youth in treatment.

Even though diagnostic status is often equated with need, these states often do not align well with people’s self-assessments overall (Pescosolido et al. 2013), but especially among racial/ethnic minority groups (Jang et al. 2014). Instead, there is considerable slippage between the presence of an underlying disorder and the perception that one needs help (Kessler et al. 2001; Moses 2009), particularly among black and Hispanic adults relative to non-Hispanic whites (Jang et al. 2014). Racial/ethnic minority populations also face a constellation of life stressors that may have unique influences on mental illness and how it is perceived, especially whether treatment is needed. For instance, exposure to everyday discrimination and residential segregation may predispose members of racial/ethnic minority groups to attribute symptoms of psychological distress to these experiences and situations and not to an underlying mental disorder. Some African Americans and Latinos, especially those with low incomes, interpret symptoms as normal responses to stressful life situations (Cabassa et al. 2008; Hines-Martin et al. 2003). In the case of Asian Americans and Latinos, there also is a tendency to report somatic complaints in the face of psychiatric disorders (Cabassa et al. 2008; Yang and Benson 2016), which may impede the realization of need for services and delay the consideration of mental health in the health care context.

**Race/Ethnicity and Perceived Need**

Some studies report that racial/ethnic minority groups are less likely than non-Latino whites to perceive a need for help (Kimerling and Baumrind 2005; Nadeem et al. 2009; Narendorf and Palmer 2016), but others report the opposite (Falck et al. 2007) or find no differences (Ault-Brutus and Alegría forthcoming; Mackenzie, Pagura, and Sareen 2010; Mojtabai et al. 2002, 2011). These discrepant findings may be due in part to methodological factors, such as sampling (e.g., select samples such as older adults) and implicit omission of some sub-populations (e.g., recent immigrants via English-only interviews). In addition, some studies examine differences for the population overall, controlling for psychiatric disorder, while other studies are limited to persons with disorders. Moreover, existing research on perceived need has almost exclusively focused on between-group differences (Edlund, Unützer, and Curran 2006; Mojtabai et al. 2002; Nadeem et al. 2009), overlooking potential heterogeneity within racial/ethnic groups. A more comprehensive examination of between- and within-group variation in perceived need with a sample representative of the population and its diversity is essential to the development of social explanations for these differences.

While past research on race/ethnicity and perceived need is inconclusive, we expect racial/ethnic differences to exist for several reasons. First, although perceptions about the causes of mental illness are generally similar across groups (Schnittker 2013), compared to non-Latino whites, a greater proportion of other racial/ethnic groups believe mental illness is caused by social stressors (Alvidrez 1999; Hines-Martin et al. 2003; Jimenez et al. 2012). Second, racial/ethnic groups may respond differently to the presence of a disorder. For example, there are notable differences in help-seeking by race/ethnicity. African Americans and Afro-Caribbeans prefer to seek out informal supports to a greater extent than non-Latino whites (Jimenez et al. 2012; Taylor et al. 2011; Woodward et al. 2010). African Americans are more likely than Caucasians to believe mental health problems will improve on their own (Anglin et al. 2008) and also are less likely to believe pharmacological treatments are acceptable or helpful (Cooper et al. 2003; Givens et al. 2007). In comparison to non-Latino whites, Asian Americans are less willing to get help from anyone (Jimenez et al., 2012), and if services are sought, counseling services are preferred over medications (Givens et al. 2007). In contrast, some Latinos in primary care are more likely than non-Latino whites to endorse a preference for medications (Jimenez et al. 2012). Yet Givens and colleagues (2007) also note that in comparison to medications, counseling is a preferred method of treatment among Latinos with depression. Lastly, attitudes do not necessarily translate into help-seeking behaviors, mirroring documented discrepancies in the application of expressed attitudes to oneself (Pescosolido et al. 2013). In particular, racial/ethnic minority groups consistently underuse mental health services (Cook et al. 2014) despite having more positive attitudes toward mental health care (Anglin et al. 2008; Schnittker, Pescosolido, and Croghan 2005) and beliefs that some forms of
treatment are beneficial to improving mental health (Givens et al. 2007; Schnittker et al. 2005)—with the notable exception of psychiatric medications for some groups (Jimenez et al. 2012; Kasckow et al. 2011; Mojtabai 2009).

Perceived need may also depend on other conjoined characteristics of race/ethnicity, such as nativity, English language proficiency (ELP), and ethnic subgroup identity (e.g., Mexican, Chinese). Although Asian Americans and Latinos constitute more than 70 percent of the U.S. immigrant population, there is considerable migration from Europe (Zong and Batalova 2015), and nearly 1 in 10 of all U.S. immigrants are black, most of whom are from Caribbean nations (Anderson 2015). White immigrant groups in particular may have better general health literacy than other groups (Sentell and Braun 2012) and therefore be more likely to engage in treatment-seeking behaviors than other racial/ethnic immigrant groups. Immigrants and those with poor ELP also tend to use mental health services infrequently, but existing studies mostly have been limited to Asian Americans and Latinos (Abe-Kim et al. 2007; Alegria et al. 2007; Bauer, Chen, and Alegria 2010). Likewise, foreign-born nativity is negatively associated with perceived need (Sareen et al. 2005), but its impact among non-Latino whites and African Americans has been largely ignored in existing research. ELP and ethnic subgroup variation in perceived need patterns are inconclusive (Bauer et al. 2012; Nguyen 2011).

**Variation by Gender and SES**

Research on race/ethnicity and perceived need tends to emphasize between-group variation while controlling for other status characteristics (e.g., Ault-Brutus and Alegria forthcoming; Nadeem et al. 2009), including gender and SES. People simultaneously belong to multiple social groups (Mays and Ghavami forthcoming), however, and the amalgamation of these statuses may impact perceived mental illness and need for treatment. Existing research has not taken into consideration how intersections of race/ethnicity with other social status categories may impact perceived need. Studies have typically assumed that multiple statuses are independent of each other, an assumption implicit in the “main effects” models used in these studies. As a result, current approaches to understanding perceived need leave unanswered questions about combinations of statuses, for instance, whether one disadvantaged status (race/ethnicity) magnifies the impact of a second one (SES), which instead requires the estimation of “conditional” or “interaction effects” models.

The possibility that individuals can be affected by multiple disadvantaged statuses has a long history that is recounted by Nguyen and colleagues (2013) for the concept of “double jeopardy” as articulated and applied to elderly black women, who are especially disadvantaged due to being black and older, followed by the development of the concepts of “triple jeopardy” and “quadruple jeopardy.” Rosenfield (2012) points to the relatively good mental health of African American women as evidence that the mental health risks associated with multiple disadvantaged statuses do not simply summate over these statuses (as implied by the main effects model and the concept of double jeopardy) but instead combine in seemingly paradoxical ways, which she attributes to self-schemas that privilege the self or others. Thus, approaches that take into account combinations of race/ethnicity, gender, and SES may be especially useful for understanding the social distribution of perceived need given that conditional relationships of this type have been documented for perceptions and attitudes about mental health care in general (Gonzalez, Alegria, and Prihoda 2005; Gonzalez et al. 2011; Ojeda and Bergstresser 2008). Furthermore, the examination of these intersections may help unmask complex relationships among multiple statuses for self-labeling in particular (Hankivsky and Grace 2015; Mays and Ghavami forthcoming).

Socioeconomic factors such as income and education—which are associated with positive attitudes toward treatment (Mojtabai 2009), better problem recognition (Thoits 2016), and more frequent help-seeking (Cook et al. 2014)—may intersect with race/ethnicity to affect perceived need. Non-Latino whites and high SES persons are more knowledgeable about mental illnesses, on average—that is, have better mental health literacy—than some racial/ethnic minority groups and low SES persons, respectively (Alvidrez 1999; Gonzalez et al. 2005; Jimenez et al. 2012). Better mental health literacy and higher economic resources are thought to affect how individuals identify and respond to their mental health problems, which may “advantage” certain groups like non-Latino whites in recognizing a need for care and seeking treatment. Hence, the benefits of educational and economic resources for the
identification of need for treatment and its response may not extend to racial/ethnic minority populations in the same ways that it benefits non-Latino whites. For example, SES is not as strongly linked to mental health literacy among African Americans and Asian Americans as whites (Lin and Cheung 1999; Thompson, Baxile, and Akbar 2004). Given these patterns, we expect SES to be positively associated with perceptions of need overall but more strongly among non-Latino whites than other racial/ethnic groups.

Similarly, it is by no means certain that gendered expectations about emotional expressiveness and help-seeking are universal across racial/ethnic groups. In general, women surpass men in perceived need (Mojtabai et al. 2002, 2011), willingness to use mental health services (Gonzalez et al. 2011), and utilization of these services (Ojeda and McGuire 2006; Wang et al. 2005), but the magnitude of these differences varies by race/ethnicity. Compared to men, traditional gender role socialization may lead women to be more expressive of feelings and engage in more help-seeking behaviors in general (Gonzalez et al. 2005), but some differences may be greater among non-Latino whites than other groups (Gonzalez et al. 2005, 2011). Similar coping strategies among African American men and women may contribute to more similar attitudes toward mental health, more so than among non-Latino whites (Rosenfield and Mouzon 2013). Additionally, adherence to Western masculine norms (e.g., emotional self-reliance, dominance) may foster negative attitudes toward mental health services and erode willingness to seek these services, particularly among Asian American men (Berger et al. 2013; Vogel et al. 2011). Yet, this tendency may be offset by higher than average levels of education among non-Latino whites and Asian Americans because education is associated with more egalitarian concepts of gender (Rosenfield and Mouzon 2013). Although existing research suggests women are more likely than men to perceive a need for treatment, we posit that this gender difference is more pronounced among non-Latino whites than other racial/ethnic groups.

The current study expands our understanding of self-labeling by examining how the intersections of race/ethnicity with gender and SES differentially influence perceptions of need for mental health care. We expect to find overall lower odds of perceived need among racial/ethnic minority groups, low SES groups, and males compared to non-Latino whites, high SES groups, and females, respectively. However, we hypothesize that the combination of these status characteristics will produce the following conditional effects: (1) High SES will be associated with high odds of perceived need for mental health care only among non-Latino whites, and (2) men will be less likely than women to perceive a need among non-Latino whites but not other groups. We also examine conjoint aspects of race/ethnicity and expect that foreign-born immigrants will have less perceived need than U.S.-born persons—and that this effect will be most prominent among Asian Americans and Latinos than other groups. Within these groups, we also anticipate heterogeneity by ELP and ethnic subgroups.

**METHODS**

Data are from the Collaborative Psychiatric Epidemiology Surveys, a national household probability sample of U.S. adults (18 years or older) comprised of three epidemiological studies: the National Comorbidity Survey-Replication (NCS-R; Kessler and Merikangas 2004), the National Survey of American Life (NSAL; Jackson et al. 2004), and the National Latino and Asian American Study (NLAAS; Alegría et al. 2004). Respondents were interviewed face-to-face using a computer-assisted structured questionnaire; a few were interviewed by telephone. Trained lay interviewers collected information about mental disorders, impairments associated with these disorders, and treatment. NSAL and NCS-R interviews were conducted in English, and NLAAS interviews were conducted in English, Spanish, Mandarin, Cantonese, Tagalog, and Vietnamese. Data collection occurred between February 2001 and November 2003.

**Sample**

The NLAAS sampled Latinos and Asian Americas (n = 4,649; 73.2 percent response rate); the NSAL sampled African Americans, Afro-Caribbeans, and non-Latino whites (n = 6,082; 70.9 percent response rate); and the NCS-R sampled all racial/ethnic groups (n = 9,282; 70.9 percent response rate). The NCS-R was completed in two parts: Part 1 included the core mental health and services measures; Part 2 included additional
psychosocial and demographic questions. Part 2 was completed by all respondents meeting diagnostic criteria for a lifetime mental disorder and a probability subsample of other respondents. The current study uses all NLAAS and NSAL participants and the Part 2 NCS-R participants (n = 5,692). When weighted, the three data sets collectively constitute a single sample that is nationally representative of the noninstitutionalized adult population of the United States (n = 16,423; for details, see Heeringa et al. 2004).

The total sample is analyzed to assess the overall association between status characteristics and perceived need in the general population, controlling for current clinical need for treatment—defined as meeting research diagnostic criteria for a 12-month mood, anxiety, and/or substance use disorder (SUD). The sample is then subdivided into: a clinical need subsample of respondents who meet lifetime diagnostic criteria for any psychiatric disorder and a no disorder subsample who do not meet these criteria because a sizeable proportion of persons who use mental health services do not have a diagnosable disorder (Druss et al. 2007).

**Measures**

**Perceived Need.** For this study, self-labeling is operationalized as perceived need for mental health care, which is a positive response to the question: “Was there ever a time during the past 12 months when you felt that you might need to see a professional because of problems with your emotions or nerves or your use of alcohol or drugs?” NSAL and NCS-R respondents who used mental health services in the past year skipped this question and were asked instead whether treatment was voluntary or due to coercion or pressure from others. Similar to previous studies using the National Comorbidity Survey or the NCS-R (Katz et al. 1997; Mojtabai et al. 2002), perceived need for these respondents was coded positive if their use was voluntary and negative if their use was coerced or pressured.

**Social Status Characteristics.** Race/ethnicity is coded: non-Latino white (the omitted reference group), Asian American (Chinese, Filipino, Vietnamese, and all other Asian origins), Latino (Puerto Rican, Cuban, Mexican, and all other Latino origins, including Afro-Caribbeans of Hispanic descent), non-Latino Afro-Caribbean, or African American. “Other race/ethnicity” is excluded due to the small sample size (n = 162) and the considerable diversity of the category. SES is operationalized as years of education in four categories (less than 12 years [reference group], 12 years, 13-15 years, and 16 or more years) and as an income-to-poverty ratio (hereafter income/poverty) calculated by dividing household income by the 2001 federal poverty threshold. Nativity is coded as foreign-born = 1, U.S.-born = 0. Aspects of ethnicity measured only in the Latino and Asian American subsamples are: ethnic subgroup (as previously listed) and ELP, which was explicitly measured in the NLAAS but assumed for NCS-R Latino and Asian American respondents because proficiency is implied by English-only interviews. For stratified analyses, nativity and ELP were combined because of sample size limitations: U.S.-born (reference group), foreign-born with fair/poor ELP, and foreign-born with excellent/good ELP.

**Clinical Need Indicators.** Clinical need consists of three dichotomous variables for 12-month research diagnosis of any mood disorder (major depression and dysthymia), anxiety disorder (generalized anxiety, agoraphobia, social phobia, panic attack, and panic disorder), or SUD (alcohol and/or drug abuse/dependence). Disorders were assessed using the World Mental Health Survey Initiative version of the Composite International Diagnostic Interview (CIDI; Kessler and Üstün 2004). Diagnostic criteria are from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association 2000). Need for care is also indexed with two measures of mental health crises: lifetime history of psychiatric hospitalizations (ever hospitalized = 1, else = 0) and lifetime suicidality (ever had suicidal ideation, plan, or attempt = 1; else = 0). In addition, self-rated mental health, an indicator of subjective need, was measured by asking, “How would you rate your overall mental health?” (1 = excellent to 5 = poor).

Need was additionally measured as nonspecific psychological distress during the past 30 days using the Kessler 6 (Kessler et al. 2003)—a six-item measure with each symptom (e.g., how often did you feel so depressed nothing could cheer your up) rated from 0 = none of the time to 4 = all of the time (range, 0-24). Functional impairment is assessed with four items from the World Health Organization’s Disability Assessment Scale
(WHODAS; WHO 2001; e.g., the number of days in the past month respondent was unable to work or carry out normal activities because of problems with their physical or mental health). Items were aggregated to create a standardized role impairment score (range, 0-100). A count of lifetime chronic physical conditions was used to take somatization into account—heart problems, hypertension, asthma, chronic lung disease, diabetes, stomach ulcers, and cancer (range, 0-7).

**Control Variables.** Insurance status was operationalized as two dichotomous variables: private insurance (i.e., employer-sponsored or privately purchased) and public insurance (i.e., Medicaid, Medicare, or other public insurance plans). Medical care from a usual place or person is coded as 1 = yes, 0 = no. There are three sociodemographic controls: age (18-29 years [omitted reference], 30-44 years, 45-59 years, and 60 or more years), employment status (employed [omitted reference], unemployed, and not in labor force), and marital status (married/cohabiting [omitted reference], previously married [divorced or widowed], and never married).

**Statistical Analysis**

Multivariable analyses were conducted using logistic regressions. The first set of analyses is for the total sample and its two subsamples defined by lifetime diagnosis status: the clinical need and no disorder subsamples (Table 2). We start by modeling the main effects of race/ethnicity, gender, SES, and nativity on perceived need. Intersections of race/ethnicity with other status characteristics were then examined by estimating conditional effects models that included two-way product interaction terms of: race/ethnicity with gender, race/ethnicity with SES, and race/ethnicity with nativity. Interactions were tested using adjusted Wald tests. Average marginal effects in predicted probabilities of statistically significant interactions are presented to aid interpretation (Norton, Wang, and Ai 2003). For the post-estimation tests, we compare all racial/ethnic groups to one another, set covariates to their means, and adjust for multiple comparisons.

The second set of analyses is limited to the clinical need subsample and stratified by race/ethnicity. Stratified analyses are presented irrespective of the significance of interaction terms in the previous analysis to ascertain whether there are pervasive differences across groups for multiple variables that go beyond a specific interaction. Part A of the stratified analyses tests factors that are shared in common across racial/ethnic groups (Table 3). For each group, the main effects model was tested against the same model with the addition of a SES by gender interaction (education by gender and income/poverty by gender). Part B tests the addition of ELP and ethnic subgroups for Latinos and Asian Americans (not shown). Stratified models were modified to address smaller sample sizes: age was continuous, education was dichotomized (less than or high school equivalent = 0, more than high school = 1), insurance and employment status were dropped due to nonsignificance, and the Afro-Caribbean subsample was dropped.

About one-quarter of respondents are missing on self-rated mental health because of a design feature of the NCS-R. We imputed this variable with ordinal logistic regression; 10 data sets were imputed using the STATA SE 14 multiple imputation chained procedures (StataCorp 2015). Respondents missing on other variables were excluded (8.3 percent missing). Rubin’s rules were used to combine results across imputed data sets (Rubin 1987).

All analyses used sample design variables and weights developed by Heeringa and colleagues (2004) to correct for stratification, clustering, unequal selection probabilities, and nonresponse. STATA SE 14 survey procedures were used to estimate all statistical models (StataCorp 2015). Standard errors were calculated using Taylor Series approximation.

**RESULTS**

**Sample Characteristics**

Table 1 presents the distribution of variables for the total sample; the clinical need subsample as defined by lifetime history of any mood, anxiety, or SUD; and the no disorder subsample. Overall, somewhat more respondents were female than male (52.4 percent vs. 47.6 percent, respectively), non-Latino whites predominated (70.0 percent), a majority had at least some college education (51.1 percent), and incomes averaged four times above the poverty line. Respondents, on average, were middle-aged. The overwhelming majority had health insurance (86.2 percent) and a usual source of care (87.3 percent). The sociodemographic profiles of the two subsamples approximate those of the total sample
except the clinical need sample has a disproportionately high representation of non-Latino whites (76.5 percent) and females (54.7 percent) and a disproportionately low representation of foreign-born persons (8.6 percent).

The two subsamples have distinctly different mental health profiles, as expected based on the criterion for selection into these subsamples. For the clinical need subsample, approximately half meet diagnostic criteria for any past-year disorder: 2 in 5 for anxiety disorders, 1 in 5 for mood disorders, and 1 in 10 for SUDs. There also is a sizeable concentration of psychiatric hospitalizations and suicidality in this subsample. Few persons in the

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample (n = 14,906)</td>
</tr>
<tr>
<td>Unweighted</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>Asian American</td>
</tr>
<tr>
<td>Latino</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
</tr>
<tr>
<td>African American</td>
</tr>
<tr>
<td>Non-Latino white</td>
</tr>
<tr>
<td>Years of education</td>
</tr>
<tr>
<td>Less than 12 years</td>
</tr>
<tr>
<td>12 years</td>
</tr>
<tr>
<td>13-15 years</td>
</tr>
<tr>
<td>16 or more years</td>
</tr>
<tr>
<td>Income-to-poverty ratio</td>
</tr>
<tr>
<td>Foreign-born</td>
</tr>
<tr>
<td>Disorder history</td>
</tr>
<tr>
<td>No lifetime disorder</td>
</tr>
<tr>
<td>Any 12-month disorder</td>
</tr>
<tr>
<td>Any lifetime disorder only</td>
</tr>
<tr>
<td>12-month mood disorder</td>
</tr>
<tr>
<td>12-month anxiety disorder</td>
</tr>
<tr>
<td>12-month substance use disorder</td>
</tr>
<tr>
<td>Lifetime hospitalization</td>
</tr>
<tr>
<td>Lifetime suicidality</td>
</tr>
<tr>
<td>Role impairment (WHODAS)</td>
</tr>
<tr>
<td>Psychological distress (K6)</td>
</tr>
<tr>
<td>Self-rated mental healthd</td>
</tr>
<tr>
<td>Uninsured</td>
</tr>
<tr>
<td>Usual source of care</td>
</tr>
<tr>
<td>Perceived need</td>
</tr>
</tbody>
</table>

Note. CPES = Collaborative Psychiatric Epidemiology Surveys; WHODAS = World Health Organization’s Disability Assessment Scale; K6 = Kessler 6.

aTotal CPES sample includes all National Latino and Asian American Study (NLAAS), National Survey of American Life (NSAL), and Part 2 National Comorbidity Survey-Replication (NCS-R) participants.
bRespondents who meet DSM-IV lifetime diagnostic criteria for any mood, anxiety, or substance use disorder.
cRespondents who do not meet lifetime diagnostic criteria for any disorder.
dSelf-rated mental health, 1 = excellent to 5 = poor.
no disorder subsample report these mental health–related crises. Individuals in the no disorder subsample also report better self-rated mental health, on average, than persons in the clinical need subsample.

About one in three persons with a clinical need report a perceived need for care. Among those with a perceived need, 17.6 percent have no history of disorder, which attests to the importance of identifying the determinants of perceived need in this group.

**Social Status and Perceived Need: Total Sample**

Table 2 presents the main and conditional effects (interaction effects) models for the total sample. For the main effects model, Asian Americans and African Americans have significantly lower odds than non-Latino whites of perceiving a need for mental health care, other factors held constant, but not Latinos and Afro-Caribbeans (Model IA). Post-estimation tests additionally reveal that relative to Latinos, Asian Americans (odds ratio [OR] = .47, \( p < .001 \)), Afro-Caribbeans (OR = .45, \( p = .04 \)), and African Americans (OR = .53, \( p < .001 \)) have lower odds of realizing a need (not shown in Table 2). Men have lower odds than women of perceiving they need help, other factors held constant. Persons with high incomes and high education versus low levels of education are more likely to perceive a need. Nativity is not statistically significant, net of other factors. Thus, there is evidence of differential self-labeling across race/ethnicity, gender, and SES insofar as these variables are statistically significant net of clinical need and other covariates.

The presence of mental health problems is the dominant influence on whether individuals think they need professional help (Model IA). Meeting 12-month diagnostic criteria is strongly associated with perceived need, somewhat more so for mood and anxiety disorders than SUDs. Lifetime suicidality, psychiatric hospitalizations, role impairments, psychological distress, and poor subjective mental health ratings are also associated with relatively high odds of perceived need. Other significant contributors to having a perceived need include having poor physical health and a usual source of care. The significant coefficient for being previously married (relative to currently married/cohabiting) may represent an impact of life problems as distinct from mental health problems per se. People who are late midlife and older have lower odds than young adults, especially those who are 60 years of age and older.

The main effects model presented assumes that the correlates of perceived need have independent effects across all racial/ethnic groups, which means the parameter estimates are constrained by default to be equivalent across groups, an assumption that may not be warranted because these groups have distinct social and economic profiles. For this sample, women represent a larger proportion of African Americans (56.1 percent) than other groups (not shown). The mode for education is less than high school for Latinos (42.6 percent), high school for African Americans (37.4 percent) and Afro-Caribbeans (30.4 percent), and college graduate or more for Asian Americans (42.9 percent); over half of non-Latino whites (55.9 percent) received at least some college education. In contrast, 10.2 percent of Latinos, 21.8 percent of Afro-Caribbeans, and 14.1 percent of African Americans completed college or pursued graduate education. Income/poverty ratios are 1.8 and 2.1 times those of African Americans for non-Latino whites and Asian Americans, respectively. Foreign-born immigrants are a majority for Asian Americans (77.2 percent), Afro-Caribbeans (66.6 percent), and Latinos (51.5 percent) but only 2.4 percent of African Americans and 3.2 percent of non-Latino whites.

We examined statistical interaction terms that enable the impact of social and economic status characteristics on perceived need to vary by race/ethnicity. Most of the two-way interactions are not statistically significant, with the exception of race/ethnicity with gender and with nativity (Model IB, Table 2). To interpret the race/ethnicity by gender interaction terms, differences in predicted probabilities of perceived need were calculated and plotted for men and women across racial/ethnic groups; all other covariates were set at their mean values (Figure 1a). Generally, men are less likely than women to believe they need treatment, but this difference is statistically significant only among non-Latino whites (\( p < .001 \)) and not among other groups. Although Figure 1a suggests similar predicted probabilities of perceived need between African American men and women (given the overlapping confidence intervals [CIs]), the logistic model results indicate that
Table 2. Logistic Regressions of Perceived Need for Mental Health Care on Social Status, Demographic Characteristics, and Need: Weighted Sample, (OR [95 Percent CI]).

<table>
<thead>
<tr>
<th></th>
<th>I. Total Sample (n = 14,906)</th>
<th>II. No Disorder Subsample (n = 8,043)</th>
<th>III. Clinical Need Subsample (n = 6,863)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. Main Effects Model</td>
<td>B. Interaction Effects Model</td>
<td>A. Main Effects Model</td>
</tr>
<tr>
<td><strong>Social status characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Latino white</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Asian American</td>
<td>.51*** [.36, .73]</td>
<td>.72 [47, 1.09]</td>
<td>.78 [39, 1.58]</td>
</tr>
<tr>
<td>Latino</td>
<td>1.09 [83, 1.41]</td>
<td>.96 [.72, 1.28]</td>
<td>1.27 [.71, 2.25]</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
<td>.49 [22, 1.07]</td>
<td>.43* [.21, .86]</td>
<td>1.02 [.40, 2.60]</td>
</tr>
<tr>
<td>African American</td>
<td>.58*** [45, .74]</td>
<td>.54*** [.42, .71]</td>
<td>.66 [.39, 1.13]</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 12 years</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>12 years</td>
<td>.93 [.70, 1.24]</td>
<td>.92 [.69, 1.23]</td>
<td>.71 [.42, 1.19]</td>
</tr>
<tr>
<td>13-15 years</td>
<td>1.20 [.89, 1.62]</td>
<td>1.18 [.88, 1.60]</td>
<td>.92 [.58, 1.48]</td>
</tr>
<tr>
<td>16+ years</td>
<td>1.76*** [1.27, 2.44]</td>
<td>1.75*** [1.26, 2.41]</td>
<td>1.41 [.87, 2.31]</td>
</tr>
<tr>
<td>Income-to-poverty ratio</td>
<td>1.03* [.101, 1.06]</td>
<td>1.03* [.101, 1.06]</td>
<td>1.03 [.98, 1.07]</td>
</tr>
<tr>
<td>Foreign-born nativity</td>
<td>.84 [.65, 1.10]</td>
<td>.15 [.64, 2.04]</td>
<td>.69 [.40, 1.19]</td>
</tr>
<tr>
<td><strong>Clinical need indicators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-month mood disorder</td>
<td>2.55*** [2.10, 3.09]</td>
<td>2.55*** [2.10, 3.10]</td>
<td>2.23*** [1.83, 2.72]</td>
</tr>
<tr>
<td>12-month anxiety disorder</td>
<td>2.61*** [2.22, 3.06]</td>
<td>2.60*** [2.22, 3.05]</td>
<td>1.90*** [1.61, 2.26]</td>
</tr>
<tr>
<td>Lifetime hospitalization</td>
<td>1.51** [1.16, 1.97]</td>
<td>1.51** [1.16, 1.97]</td>
<td>1.51 [81, 2.80]</td>
</tr>
<tr>
<td>Role impairment score</td>
<td>1.00* [100, 1.01]</td>
<td>1.00* [100, 1.01]</td>
<td>1.01* [100, 1.02]</td>
</tr>
<tr>
<td>Number of chronic conditions</td>
<td>1.19*** [1.06, 1.34]</td>
<td>1.19** [1.06, 1.34]</td>
<td>1.23* [1.00, 1.52]</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurance status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninsured</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Privately insured</td>
<td>1.08 [.89, 1.30]</td>
<td>1.07 [.89, 1.29]</td>
<td>1.25 [.78, 2.01]</td>
</tr>
<tr>
<td>Publicly insured</td>
<td>1.14 [.90, 1.43]</td>
<td>1.14 [.91, 1.44]</td>
<td>1.34 [.73, 2.46]</td>
</tr>
<tr>
<td>Usual source of care</td>
<td>1.39* [.107, 1.82]</td>
<td>1.40* [.107, 1.83]</td>
<td>1.78* [1.02, 3.08]</td>
</tr>
<tr>
<td>Age</td>
<td>I. Total Sample (n = 14,906)</td>
<td>II. No Disorder Subsample (n = 8,043)</td>
<td>III. Clinical Need Subsample (n = 6,863)</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td></td>
<td>A. Main Effects Model</td>
<td>B. Interaction Effects Model</td>
<td>A. Main Effects Model</td>
</tr>
<tr>
<td>18-29 years old</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>30-44 years old</td>
<td>.98 [.79, 1.20]</td>
<td>.98 [1.0, 1.21]</td>
<td>.96 [1.59]</td>
</tr>
<tr>
<td>45-59 years old</td>
<td>.74*** [.62, .88]</td>
<td>.74** [.62, .89]</td>
<td>.75 [1.44, 1.26]</td>
</tr>
<tr>
<td>60 years old and over</td>
<td>.28*** [.18, .43]</td>
<td>.28*** [.18, .42]</td>
<td>.17** [.05, .52]</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Previously married</td>
<td>1.37** [1.08, 1.74]</td>
<td>1.36** [1.08, 1.72]</td>
<td>1.27 [1.71, 2.27]</td>
</tr>
<tr>
<td>Never married</td>
<td>1.08 [1.88, 1.32]</td>
<td>1.07 [1.87, 1.31]</td>
<td>1.31 [1.76, 2.27]</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.03 [1.77, 1.37]</td>
<td>1.03 [1.77, 1.37]</td>
<td>1.23 [1.68, 2.23]</td>
</tr>
<tr>
<td>Gender by race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Latino white, male</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Asian, male</td>
<td>1.53* [1.00, 2.35]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latino, male</td>
<td>1.46* [1.09, 1.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afro-Caribbean, male</td>
<td>1.38** [26, 7.48]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American, male</td>
<td>1.38 [1.91, 1.65]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nativity by race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Latino white, foreign</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Asian American, foreign</td>
<td>.36** [.18, .71]</td>
<td></td>
<td>.23*** [.10, .52]</td>
</tr>
<tr>
<td>Latino, foreign</td>
<td>.69 [1.35, 1.36]</td>
<td></td>
<td>.62 [28, 1.33]</td>
</tr>
<tr>
<td>Afro-Caribbean, foreign</td>
<td>.76 [22, 2.54]</td>
<td></td>
<td>.52 [12, 2.29]</td>
</tr>
<tr>
<td>African American, foreign</td>
<td>.60 [25, 2.48]</td>
<td></td>
<td>.24 [.05, 1.06]</td>
</tr>
<tr>
<td>Test of gender by race/ethnicity</td>
<td>p = .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of foreign-born by race/ethnicity</td>
<td>p = .02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. OR = odds ratio; CI = confidence interval.

*p ≤ .05. **p ≤ .01. ***p ≤ .001.

Self-rated mental health scores range from 1 = excellent to 5 = poor.
this gender difference is statistically significant for this group \( (p < .01) \). This inconsistency may arise because predicted probabilities are sensitive to the values set for the covariates, while the odds ratios are not. We give priority to the results of the logistic model for this reason and because the predicted probabilities are derived from this statistical model. The gender difference is also significant among the African American clinical need subsample in the stratified analysis (see Table 3). In this instance, the coefficient for gender is estimated with all other independent variables estimated separately among African Americans with a clinical need; by contrast, Model IB (Table 2) estimates the gender effect holding the effects of all other independent variables (except nativity) constant across racial ethnic groups in the total sample. Females have significantly higher levels of perceived need than males among African Americans with clinical need. Therefore, we conclude that there are gender differences among African Americans overall and for the clinical need subsample. In multiple comparison tests (not shown), no other differences between racial/ethnic groups were observed. Correspondingly, racial/ethnic group differences in perceived need in the total sample are somewhat greater among women than men.

Figure 1b plots the predicted probabilities of perceiving a need for mental health care by nativity and race/ethnicity: Foreign-born Asian Americans are less likely to perceive a need than U.S.-born Asian Americans, and this nativity difference is statistically significant \( (p < .001) \). Additionally, this nativity difference is largest among Asian Americans when compared to non-Latino whites; however, this difference disappears when adjusted for multiple comparisons (not shown), and no other group differences were observed. In sum, the significant interactions in the total sample suggest that gender and nativity differences in self-labeling are conditional on race/ethnicity.

**Social Status and Perceived Need: Clinical Need Subsample**

Model IIIA in Table 2 presents the main effects for the subsample who meets research diagnostic criteria for any lifetime mood, anxiety, and/or SUD. With the exception of Latinos, racial/ethnic minority groups have lower odds of perceiving a need for care than non-Latino whites, other variables held constant. In post-estimation tests of other contrasts (not shown), we also find that Asian Americans \( (OR = .44, p < .001) \), Afro-Caribbeans \( (OR = .37, p < .01) \), and African Americans \( (OR = .54, p < .001) \) have significantly lower odds of perceived need than Latinos. No other group comparisons are statistically significant. Men have significantly lower odds than women. Other things being equal, SES has a limited impact, with the only significant difference being the higher odds among people with at least a college education compared to less than a high school education. Foreign-born nativity is not significant.

All indicators of clinical need for services are significantly associated with greater odds of perceived need for professional help, with the exception of role impairments (Model IIIA). Compared to those with no past-year disorder, the odds of seeing oneself as needing care are roughly two times as great for persons with a mood, anxiety, or substance use disorder. Other things being equal, these odds are also greater among those

**Social Status and Perceived Need: No Disorder Subsample**

Persons who do not meet criteria for a lifetime diagnosis are significantly less likely to see themselves as needing professional help than persons who meet these criteria \( (6.0 \text{ percent vs. } 33.3 \text{ percent, Table 1}) \). Social status characteristics have a limited impact on self-labeling in the no disorder subsample (Model IIA, Table 2). Only one status characteristic is statistically significant in this subsample. Males have lower odds of perceived need than females, other factors held constant. Furthermore, there are no significant interactions among status characteristics.

Suicidality, role impairments, and nonspecific psychological distress are significantly associated with larger odds of perceived need; lifetime psychiatric hospitalization is not, which may be due to its extremely low occurrence in this subsample (1.7 percent). Self-rated mental health and physical health status as assessed by the number of chronic conditions also are significant in the expected direction. Similar to the total sample, age differences are found—those in the oldest age group have lower odds than those in the youngest group, and having a usual source of care is associated with relatively high odds of seeing oneself as needing mental health care.
who are positive for lifetime psychiatric hospitalizations and suicidality than among those who are negative on these variables. Psychological distress, poor self-rated mental health, and chronic conditions are associated with greater odds of perceiving a need. The odds of seeing oneself as needing help are lower among persons who are late middle age and older, compared to young adults. Previously married persons have greater odds than married persons.

Turning to the conditional model, only one statistically significant interaction term is found of the four tested: race/ethnicity with nativity (Model IIIB). Figure 2 plots the predicted probabilities of

**Figure 1.** Predicted probabilities of perceiving a need for mental health care across (a) gender and race/ethnicity and (b) nativity and race/ethnicity: total sample.
perceiving a need for mental health care across nativity and race/ethnicity for the clinical need subsample. Similar to the total sample, the probability of perceived need is significantly lower among foreign-born than U.S.-born Asian Americans \( (p < .001) \). This nativity effect is statistically significant only among Asian Americans. The nativity difference among this group is also largest when compared to non-Latino whites. There are no significant nativity differences for other racial/ethnic groups.

**Variation in Perceived Need within Racial/Ethnic Groups**

The first set of stratified analyses assessing within-group heterogeneity in perceived need examines factors that are relevant to all racial/ethnic groups with a clinical need (Table 3). Social status characteristics have some impact on perceived need within all groups, net of other factors. Males are significantly less likely than women to believe they need mental health services among African Americans and non-Latino whites, but gender is not significant among Asian Americans or Latinos. Likewise, high education is associated with higher odds of perceived need for African Americans and non-Latino whites only. Lastly, foreign-born nativity is significantly associated with lower odds among Asian Americans only.

We also examined potential interactions between gender and SES within each group; all but one were statistically nonsignificant. Among African Americans with a clinical need, a statistically significant interaction between gender and income/poverty was found (not shown). Further investigation revealed that the effect of income/poverty on perceived need was positive but statistically nonsignificant among African American women and negative but only marginally significant among African American men \( (p = .054) \).

For the most part, diagnostic status indicators (with the exception of past-year SUD in most stratified models), psychological distress, and self-rated mental health are strong and fairly consistent correlates of perceived need within groups. Lifetime hospitalization is associated with higher odds of perceived need for Asian Americans and Latinos but not for African Americans or non-Latino whites. Suicidality is associated with high odds among all groups except Asian Americans, whereas role impairments are associated with higher odds of perceived need among Asian Americans only. With respect to other covariates, older Latinos and non-Latino whites are less likely to perceive a need for care than younger persons. Usual care has a particularly large impact among Asian Americans. Lastly, compared to married/cohabiting adults, non-Latino whites who are previously married are more likely to self-label as having a need for care, while single African Americans are less likely to perceive a need.

The second set of stratified models of Asian Americans and Latinos examined characteristics assessed only for these groups—ELP and ethnic subgroup identity (not shown). Ethnic subgroup variation in perceived need was not found within either the Asian American or the Latino groups, net of other variables in the model. Compared to U.S.-born Asian Americans, foreign-born Asian Americans overall, irrespective of ELP, are less likely to perceive a need for mental health treatment. Foreign-born nativity and ELP are not significant for Latinos.

**DISCUSSION**

The current study examines how the intersections of race/ethnicity with gender and SES are associated with self-labeling, defined as perceiving that one needs treatment for mental health–related problems. Perceived need commands our attention because it plays a pivotal role in voluntary help-seeking and may contribute to mental health care disparities. A key determinant of perceived need is clinical need, indexed in this study as psychiatric disorder, nonspecific psychological distress, and social role impairments. Variation in perceived need by social status characteristics that persist net of these indicators points to potential disparities that may be passed along to the use of services, particularly among persons with psychiatric disorders. Therefore, it is imperative to have an accurate accounting of the associations between these status characteristics and perceived need.

We hypothesized that race/ethnicity combines with gender, SES, and the related characteristic of nativity to influence perceived need. Specifically, we expected the following conditional effects: (1) high SES groups to have greater perceived need than low SES groups but only among non-Latino whites, (2) men to be less likely than
Table 3. Logistic Regressions of Perceived Need for Mental Health Care on Social Status, Demographic Characteristics, and Need: Clinical Need Subsample Stratified by Race/Ethnicity (OR [95 Percent CI]).

<table>
<thead>
<tr>
<th></th>
<th>Asian American (n = 538)</th>
<th>Latino (n = 1,359)</th>
<th>African American (n = 1,745)</th>
<th>Non-Latino White (n = 2,789)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social status characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.75 [0.35, 1.63]</td>
<td>.75 [0.54, 1.05]</td>
<td>.68* [0.50, .92]</td>
<td>.57*** [0.49, .68]</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Male</td>
<td>1.55 [0.73, 3.32]</td>
<td>1.21 [0.88, 1.68]</td>
<td>1.99*** [1.52, 2.60]</td>
<td>1.54** [1.14, 2.07]</td>
</tr>
<tr>
<td>Male</td>
<td>1.04 [0.99, 1.09]</td>
<td>1.03 [0.98, 1.08]</td>
<td>0.98 [0.94, 1.04]</td>
<td>1.04* [1.00, 1.07]</td>
</tr>
<tr>
<td>Male</td>
<td>0.40** [0.20, 0.78]</td>
<td>0.88 [0.63, 1.23]</td>
<td>0.43 [0.24, 1.26]</td>
<td>1.60 [0.81, 3.19]</td>
</tr>
<tr>
<td>Clinical need indicators</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4.43*** [2.25, 8.73]</td>
<td>2.20*** [1.54, 3.15]</td>
<td>2.69*** [1.83, 3.95]</td>
<td>2.21*** [1.70, 2.86]</td>
</tr>
<tr>
<td>Male</td>
<td>2.96 [0.51, 17.28]</td>
<td>2.19 [0.74, 6.48]</td>
<td>1.96* [1.02, 3.76]</td>
<td>1.46 [0.91, 2.35]</td>
</tr>
<tr>
<td>Male</td>
<td>5.15*** [1.94, 13.65]</td>
<td>2.22*** [1.42, 3.48]</td>
<td>1.25 [0.80, 1.94]</td>
<td>1.23 [0.88, 1.73]</td>
</tr>
<tr>
<td>Male</td>
<td>1.81 [0.99, 3.31]</td>
<td>1.49* [1.08, 2.05]</td>
<td>1.79*** [1.28, 2.49]</td>
<td>1.38*** [1.17, 1.63]</td>
</tr>
<tr>
<td>Male</td>
<td>1.01* [1.00, 1.02]</td>
<td>1.00 [0.99, 1.00]</td>
<td>1.00 [0.99, 1.01]</td>
<td>1.00 [0.99, 1.01]</td>
</tr>
<tr>
<td>Male</td>
<td>.99 [0.90, 1.10]</td>
<td>1.10*** [1.06, 1.15]</td>
<td>1.06** [1.02, 1.10]</td>
<td>1.12*** [1.09, 1.16]</td>
</tr>
<tr>
<td>Male</td>
<td>1.48* [1.05, 2.07]</td>
<td>1.42*** [1.21, 1.67]</td>
<td>1.29** [1.09, 1.54]</td>
<td>1.38*** [1.15, 1.66]</td>
</tr>
<tr>
<td>Male</td>
<td>1.00 [0.75, 1.33]</td>
<td>1.19 [0.98, 1.45]</td>
<td>1.13 [0.98, 1.30]</td>
<td>1.16 [0.98, 1.38]</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual source of care</td>
<td>2.57* [1.17, 5.63]</td>
<td>1.24 [0.80, 1.90]</td>
<td>1.36 [0.86, 2.14]</td>
<td>1.34 [0.94, 1.92]</td>
</tr>
<tr>
<td>Age</td>
<td>1.00 [0.98, 1.03]</td>
<td>0.99* [0.97, .99]</td>
<td>0.99 [0.98, 1.01]</td>
<td>0.98*** [0.97, .99]</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>.75 [0.26, 2.20]</td>
<td>.94 [0.54, 1.64]</td>
<td>.91 [0.68, 1.22]</td>
<td>1.52** [1.14, 2.03]</td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>1.31 [0.56, 3.10]</td>
<td>0.62 [0.34, 1.13]</td>
<td>0.67* [0.46, .98]</td>
<td>0.98 [0.74, 1.30]</td>
</tr>
</tbody>
</table>

Note. OR = odds ratio; CI = confidence interval.
*p ≤ .05. **p ≤ .01. ***p ≤ .001.
women to believe a need for care but only among non-Latino whites, and (c) foreign-born nativity to have a negative effect on perceived need among Asian American and Latino groups but not among other groups. Our findings provide some support for these conditional effects in the form of statistically significant interactions: two for the full sample—between race/ethnicity and gender and race/ethnicity and nativity—and one for the clinical need subsample between race/ethnicity and nativity. The impact of higher education and income on perceived need is not conditional on race/ethnicity for the total sample or the two subsamples, an unanticipated result (however, we note that an income by gender interaction was found in the African American clinical need subsample). The alternative, that each status has the same impact on perceived need irrespective of a person’s other status characteristics, operationalized as the main effects model, is preferred for the no psychiatric disorder subsample. In other words, there was no evidence of conditional effects in self-labeling for the no disorder subsample. Finally, quite a few interactions were tested, and only four were statistically significant, which means that the findings overall tilt in favor of independent effects of social status characteristics on self-labeling rather than conditional effects.

Self-labeling theory posits that socially and economically disadvantaged persons are less likely than advantaged persons to apply mental illness–related labels to themselves and engage in voluntary help-seeking. The results from the main effects models provide some support for this assertion in the total sample and clinical need subsample (with the caveat that the corresponding conditional model is the preferred model). Net of clinical need and controls, the odds of having a perceived need are lower among racial/ethnic minority groups relative to non-Latino whites, with the exception of Latinos, and among those with the lowest level of education relative to the highest. Higher education may facilitate the recognition of mental health problems by improving mental health literacy.

Furthermore, self-labeling theory implies that the person who is most likely to self-label is a well socialized actor who is experiencing abnormal states and endorses dominant cultural perspectives toward mental health and treatment, which in contemporary American society favors the medical model of mental illness. From this perspective, the low perceived need among Asian Americans, Afro-Caribbeans (clinical need subsample only), and African Americans may stem from cultural differences in beliefs about mental illness and its

Figure 2. Predicted probabilities of perceiving a need for mental health care across nativity and race/ethnicity: clinical need subsample.
etiology as well as variations in culturally acceptable ways to respond to mental illness. The similarity in perceived need between Latinos and non-Latino whites may result from comparable or more favorable attitudes toward help-seeking and mental health services observed among these groups (Gonzalez et al. 2005; Mojtabai 2007).

In contrast, the observed gender difference contradicts this theory because women, who on average are socially and economically disadvantaged relative to men, are more likely than men to think they need care. The most likely influence for the gender exception is the continued existence of traditional gender role prohibitions against the expression of emotions and help-seeking by males. Based on the main effects models, we conclude that social disadvantage is not the sole factor at work in self-labeling as it pertains to perceived need for services; some other factors appear to have the capacity to offset at least some specific manifestations of disadvantage.

However, the conditional model is preferred over the main effects model for the total sample and the clinical need subsample because differences in perceived need by gender and nativity are contingent on race/ethnicity, as indicated by the significant interactions among these variables. The first interaction applies to the total sample: Overall, men have significantly lower odds than women of thinking they need help, but the magnitude of this difference varies by race/ethnicity and is only statistically significant among non-Latino whites and African Americans. We expected to find significant gender differences in non-Latino whites, but the significant difference among African Americans was unforeseen. Although masculine norms are associated with less positive attitudes toward mental health treatments, this effect seems to be particularly strong among African American men compared to European American men (Vogel et al. 2011), which may contribute to the observed low perceived need among African American men. Additionally, non-Latino white men are less likely to self-label than other women, particularly non-Latino white and Latina women, perhaps as a way to avoid stigma because they may think they have more to “lose” by adopting a mental illness label, thereby jeopardizing their dominant social status (Ojeda and Bergstresser 2008). The similarity between men and women among Asian Americans and Latinos may result from cultural norms about keeping personal and family troubles private, which may discourage both men and women from even thinking about turning toward professionals for help. This interpretation is consistent with research showing that men have significantly more negative attitudes toward mental health services than women among non-Latino whites but not Latinos (Gonzalez et al. 2005).

The second interaction applies to both the total sample and clinical need subsample: Differences in perceived need by nativity are greater among Asian Americans than other groups, in part consistent with our hypotheses; the lack of a nativity effect among Latinos was not expected. Foreign-born Asian Americans are less likely to perceive a need for mental health care than U.S.-born Asian Americans, and this difference is unexpectedly not affected by English language proficiency as evidenced in the stratified analyses. This nativity effect is consistent with reports of low utilization of mental health services among immigrant Asian Americans (Abe-Kim et al. 2007). Poorer self-labeling among Asian American immigrants compared to their U.S.-born counterparts may reflect poorer health literacy and more stringent cultural norms against disclosing private information to outsiders within immigrant subpopulations (Leong and Lau 2001). Aside from Latinos, the lack of an effect of nativity for the other groups may be a function of low rates of immigration as evidenced by the wide confidence intervals. Lastly, we did not find support for our hypothesis that other conjoined aspects of ethnicity (ELP, ethnic subgroup identity) would impact perceived need for Latinos and Asian Americans.

For the sample as a whole, our results suggest that perceiving a need for mental health care is a complex function of different dimensions of social stratification that transcend the relationships specified by self-labeling theory. Social groups that have historically experienced cumulative social and economic disadvantage theoretically should be at highest risk of not self-labeling. We observe these differences in the main effects of race/ethnicity and SES. However, the preferred conditional effects model tells a different story: Social disadvantage does not always translate to poor self-labeling given paradoxical patterns in perceived need across gender and race/ethnicity. Despite their advantaged position in society, non-Latino white men have a low predicted probability of perceived need whereas these probabilities are high among non-Latino white women and Latinos irrespective of gender. Social and
economic disadvantage do not appear to be sufficient alone to account for these patterns; instead, variation in gender roles across racial/ethnic groups appear to be implicated. The idea that one disadvantaged status magnifies the impact of another disadvantaged status is not supported insofar as there were no statistically significant synergistic interactions.

Heterogeneity in the impact of SES, principally education, is seen in the within racial/ethnic groups stratified analyses for the clinical need subsample. Education is significantly associated with perceived need but only among African Americans and non-Latino whites. Higher education is thought to convey greater knowledge about mental illnesses and their treatment, knowledge that may become relevant to perceived need primarily when people experience firsthand signs and symptoms that are recognizable as psychiatric disorders. An interaction between gender and income/poverty also was observed within the African American clinical need subsample (not shown), with income/poverty having a marginally significant negative effect on perceived need for men and not women, contrary to self-labeling theory. The high proportion of immigrant Asian Americans and Latinos may account for the absence of an education effect in these groups if strong cultural influences among immigrants offset the impact of education and income among non-immigrants.

Among persons with no disorders, only one status characteristic is associated with perceived need: Men have lower odds than women, taking into consideration other status and demographic characteristics, indicators of clinical need (other than diagnosable disorder), and enabling factors (health insurance and usual source of care). Of the relevant indicators of clinical need included in this model, all are significantly associated with perceived need except lifetime psychiatric hospitalization, which is quite rare in this subsample and may have limited power. That perceived need is associated with clinical factors indicates that perceptions are grounded in genuine experiences of mental health problems, which resurrects a long-standing debate in sociology over the reification of diagnosis (Mirowsky and Ross 1989). In an ideal world, perceived need should be independent of social status when actual need is taken into account; therefore, the results for the no disorder subsample approach an absence of health disparities.

Several limitations merit mention. First, some variables could not be evaluated because they were not measured, including the impact of stigma on perceived need, which may be a source of omitted variable bias. Likewise, discrimination can negatively affect mental health treatment among racial/ethnic minority groups (Mays et al. 2017), but its impact on perceived need could not be assessed for the total sample due to extensive data missing by design. Provisional analysis suggests that discrimination is associated with perceived need only among African Americans and not Latinos or Asian Americans. Future research should more thoroughly examine this association, including discrimination experienced within the health care system. These assessments should be conducted for the entire population and not exclude non-Latino whites because all groups contain people who are at risk for discrimination on some traits and because discrimination based on mental illness and treatment is relevant to all racial/ethnic groups.

Second, the NCS-R and NSAL did not measure perceived need directly for all respondents but instead inferred it for respondents who used mental health services in the past 12-months and then reported that the use was voluntary; these persons skipped out of the perceived need question. This procedure is logical, but it is based on a rational choice model, whereas the process leading to treatment is often haphazard. People who “muddle through” are unsure about whether treatment was entirely voluntary or not (Pescosolido, Gardner, and Lubell 1998); thus, reports of “voluntary” treatment may nevertheless reflect subtle coercion or pressure by others, but the data set does not contain measures to discern these subtleties. Yet again, the direct assessments of perceived need also are subjective assessments, mitigating this concern somewhat. That having been said, perceived need has been measured this way in several published studies using the National Comorbidity Survey and the NCS-R (Katz et al. 1997; Mojtabai et al. 2002), which supports this approach. Nevertheless, future studies should assess perceived need for all respondents prior to assessing treatment.

Third, perceived need was measured for any formal mental health provider, but racial/ethnic differences in this perception may be more disparate for some types of providers (e.g., medications) than others (e.g., medical doctors). We recommend that future research in this area directly assess perceived need for specific providers.
Lastly, clinical need was defined for three classes of mental disorders that were assessed across the three CPES data sets, meaning that there is misclassification for people who met criteria for other disorders. For example, the NLAAS did not assess bipolar disorder, unlike the NCS-R and NSAL. Consequently, mood disorders do not include bipolar disorder. Likewise, psychotic disorders were either not assessed in specific surveys or generated substantial missing data. For these reasons, the association between disorder status and perceived need may be biased. However, in sensitivity analyses of respondents with data about psychotic symptoms, there was no significant association with perceived need.

These limitations notwithstanding, the current study has a number of notable strengths, including the large probability sample that permits generalization to the U.S. adult population. The use of multiple languages means that the sample more fully represents the Asian American and Latino populations than other similar studies. The large and diverse sample also enables the examination of intersections among statuses and within-group heterogeneity, which is overlooked in existing research on perceived need. Finally, our use of perceived need as an indicator of self-labeling expands previous studies that have equated self-labeling with accounts of voluntary help-seeking (Thoits 2005) or perceptions of illness identity (Moses 2009; Thoits 2016). The concept of perceived need aligns with self-labeling processes because it represents a pivotal antecedent that drives voluntary help-seeking—that is, the recognition that a problem requires professional intervention.

The findings from our study have theoretical implications toward the expansion of self-labeling theory. We utilized an intersectional approach to understand the social distribution of perceived need and found that intersections of race/ethnicity with other social status characteristics, particularly gender and nativity, matter to self-labeling. Incorporating an intersectionality framework into self-labeling theory may better help inform our understanding of the mechanisms and social processes that lead to differential self-labeling among socially and economically disadvantaged groups. This framework asserts that the intersection of different social statuses reflects interlocking structural systems of power, privilege, and disadvantage and creates unique life experiences (e.g., discrimination, social stereotypes, etc.) that impact outcomes (Crenshaw 1989, 1991; Ghavami, Katsiaficas, and Rogers 2016; Mays and Ghavami forthcoming). The application of this framework would emphasize how major social status categories and context come together to create experiences that are “more than the sum of statuses” (Crenshaw 1989, 1991; Mays and Ghavami forthcoming).

For example, racial/ethnic minority groups with high SES may not self-label as readily as high SES, non-Latino whites because of bias and discrimination within formal systems of care, which not only produces perceptions of distrust in these systems (Dovidio et al. 2008) but also disrupts treatment (Mays et al. 2017). Bauer (2014) warns, however, that imposing an intersectional approach that solely intersects social identities or positions is insufficient and recommends that these approaches also consider the examination of underlying intersectional social processes (e.g., racism, classism). Thus, adapting an intersectionality framework to self-labeling theory would not only be important for understanding self-labeling patterns across social strata but would also reveal how various statuses and social processes associated with self-labeling may lead to disparities in mental health care utilization.

The application of an intersectionality framework to self-labeling theory raises methodological considerations. The current study examines intersections with cross-product interactions, revealing whether the effect of SES on perceived need, for example, varies significantly across racial/ethnic groups—that is, whether the impact of one status differs across another. Other approaches may focus on assessing whether self-labeling varies at different intersections by combining highly correlated social status categories into a single variable and then comparing combined groups (Hankivsky and Grace 2015). For example, merging race/ethnicity and SES as one variable would allow the comparison of perceived need between high SES non-Latino whites to low SES non-Latino whites, low SES and high SES Asian Americans, and so on. This approach implies a “dissection of individuals into their stratified component identities, positions, or experiences” (Bauer 2014:15). While our study uses the former approach (i.e., examined interactions), it is the first study, to our knowledge, to provide evidence of conditional patterns in self-labeling across race/ethnicity, gender, and nativity. We recommend that future studies on self-labeling utilize alternative statistical specifications to examine how intersections of social statuses influence differential self-labeling.
Our findings suggest practical steps to improving mental health care utilization in the United States. As expected, worse mental health—as indexed by the presence of a recent psychiatric disorder, psychiatric hospitalizations, and so forth—consistently increase the odds of reporting a need for services across social statuses and subsamples. These findings indicate that actual need for services is the primary contributor to perceived need and is shared in common across diverse groups. Therefore, improving people’s knowledge of the signs and symptoms of psychiatric disorders may facilitate more accurate self-labeling and assessments of need for treatment. Furthermore, treatment of physical health conditions may function as a gateway to mental health service use insofar as chronic conditions and having a usual source of care are associated with seeing oneself as needing help for mental problems. It seems likely that persons routinely gain knowledge about mental disorders and their treatment at these sites, perhaps by consulting their physicians or other health professionals about problems they experience. Therefore, primary care settings may be important locations for programmatic efforts to improve people’s ability to recognize mental health problems and consider treatment as needed.

Self-labeling processes are essential for the voluntary use of mental health services among persons with psychiatric problems. The current study provides some support to the idea that the combination of multiple social status characteristics—including race/ethnicity with gender, nativity, and SES—may create barriers for specific groups for the recognition of mental health problems and realization that treatment is needed. Intervention efforts that target racial/ethnic groups with low rates of perceived need will need to address the unique experiences of immigrants, low SES groups, and men in order to reduce disparities in mental health services utilization. More importantly, identifying the statuses and mechanisms that lead to differential self-labeling is essential to explaining why disparities in mental health care utilization exist.

AUTHORS’ NOTE
The content is solely the responsibility of the authors and does not necessarily represent the official views of National Institutes of Health and National Institute of Mental Health.

FUNDING
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by: the National Institutes of Health (NIH)/National Center for Advancing Translational Science (NCATS) UCLA Clinical and Translation Science Institute (CTSI) grant (TL1 TR000121), a National Institute of Mental Health (NIMH) T32 postdoctoral grant (5-T32-MH 13043), and a NIH National Institute of Minority and Minority Health Disparities Grant (MD 006923).

REFERENCES


Bauer, Amy M., Chih-Nan Chen, and Margarita Alegría. 2010. “English Language Proficiency and Mental Health Service Use among Latino and Asian Americans with Mental Disorders.” Medical Care 48(12): 1097-104.


by Age, Gender, Ethnicity/Race and Education.” Social Psychiatry and Psychiatric Epidemiology 46:45-57.


Neighbors, Harold W., Cleopatra Caldwell, David R. Williams, Randolph Nesse, Robert Joseph Taylor, Kai McKeever Bullard, Myriam Torres, and James S. Jackson. 2007. “Race, Ethnicity, and the Use of Services for Mental Disorders: Results from the National Survey of American Life.” Arch Gen Psychiatry 64:485-94.
StataCorp. 2015. “Stata Statistical Software: Release 14.” College Station, TX: StataCorp LP.
General Topics to Address in Breakout Session

The following should be considered as guiding questions for the breakout sessions, not a rigid agenda to be followed.

Working from a life course perspective:

1. What are the most pressing research questions going forward? Please direct attention to racial, ethnic, and other social factors associated with mental health disparities over the life course.
   - In addition to identifying key risk factors for women’s mental health across the life course, identify important protective factors (those known and those known to be explored) that may foster resilience in mental health, especially for specific populations of women (e.g., justice-involved, veterans, substance-using, etc.)
   - Address data, methods, and research gaps

2. How do we ensure that those research questions are addressed?
   - Identify cutting-edge and forward-thinking ideas for data collection and analysis.
   - How can we foster research that addresses multiple and intersecting levels of analysis (e.g., biological, psychological, social, behavioral risk and protective factors that may vary by race, ethnicity, and other social factors, across the life course)?
   - What data collection and research strategies would be most helpful in augmenting the scientific knowledge base and developing effective interventions to address race/ethnic disparities in women’s mental health?

3. Going forward, what are the most critical practice and policy issues to address concerning women’s mental health over the life course? Consider disparities, risk factors, protective factors, and population patterns of mental health during specific periods of the life course and accumulating over the life course.
   - Identify cutting-edge and forward-thinking ideas for practice and policy.
   - Based on current research and projected trends in women’s mental health, can you make specific recommendations to guide policy and practice concerning women’s mental health over the life course?
**Group Reports Back to the Larger Group**

Breakout group discussion leaders should report back to the larger group on:

A. **What we need to know going forward**
   1. **Research needs:**
      a. What we need to know about women’s mental health going forward, gaps in data and research
         i. Attention to racial, ethnic, and other social risk factors and protective factors
   2. **Policy and practice needs:**
      a. What we need to know about women’s mental health going forward, policy and practice needs
         i. Attention to racial, ethnic, and other social risk factors and protective factors

B. **Recommendations: What we need to do going forward** (most promising and innovative directions; specific recommendations)
   a. Research
   b. Policy
   c. Practice