
October 9th, 2017.
Hartford Center of Gerontological Excellence
Featured Research on Aging forum

Presented by: Elena M. Andresen, PhD. Professor of Epidemiology; OHSU-PSU School of Public Health.

Executive Vice President and Provost; Baird Hall, 1011j. Phone 503-494-4460 email: andresee@ohsu.edu
Learning Objectives

• Become familiar with field research methods & measures in a successful metropolitan cohort
• Identify advantages & disadvantages of individual & environment (neighborhood) risk factor measures
Colleagues: NIH African American Health Cohort Study (AAH)

PI: Doug Miller, MD  
(Ret) Indiana University

Frederic Wolinsky, PhD  
University of Iowa

Ted Malmstrom, PhD  
Saint Louis University

J. Philip Miller, AB  
Washington University

Mario Schootman, PhD  
St Louis University
Background

• Health Disparity in the US
  African American seniors have more disability & greater health risks than whites

• Health Diversity in the US
  Among African Americans: substantial variability in disability levels & risks

  *Miller et al: AAH* clinically relevant depressive symptoms 21%.
  Range 14% (suburban men) to 30% (inner city women)
Two Catchment Areas for AAH

Cohort N = 998
Ages 50-64 @ baseline
50-50 geographic split

Inner city
Mississippi River
Suburbs
AAH study

Door-to-door *random sample selection* of African American adults 50-64 from two neighborhoods (*strata*) in 2000 (75% response)

1. Inner city St. Louis neighborhoods.  
   *47% annual incomes < $10,000*

2. Suburban, integrated neighborhoods.  
   *15% annual incomes < $10,000*

Completed sample 1,000 in 9+ months
AAH Cohort Study Protocol

Years 1, 4, & 10. Extensive in-home assessment & Observer ratings of neighborhood

Baseline Year 1: 2000; Follow-up 10 (year 11): early 2011

AAH Research Team Year 4
AAH Baseline Data Collection

• 75 minute in-home interview & assessment
  – Substantial functional testing, e.g., lung function, balance, gait, hand strength, weight & body composition
• Retest of key variables (reliability)
• Blood samples (biomedical hypotheses)
• Muscle strength (in-lab) tests
AAH Study Data Collection

• Monetary incentives
  – Increasing incentives, & by number of study procedures, e.g., in home (all subjects), add blood draw &/or in-lab visit, etc. ($100+)

• Neighborhood observer rating
  – Baseline: 5 item scale (*Andresen et al.*, 2005)
  – Wave 4: 20 item scale (*Andresen et al.*, 2008)
  – Wave 10: both; 27 items (*Andresen et al.*, 2013)

• Data collected at intervening annual “waves” by telephone
AAH Data Collection/Training

• One week training for in-home phases
• Includes study procedures, software/tracking procedures for computer assisted interviewing
• Interviews with standardized components (e.g., fear of falling; depressive symptoms)
• Physiologic measures (equipment & performance tests)
• Special procedures, e.g., falls safety certification for performance tests
• Neighborhood (observer) rating
Wave 4 interviewer training

Timed 4-meter walk

Timed tandem stand
AAH Study Cohort Maintenance

• Periodic newsletters, health hints
  – E.g., tips about heat waves, food safety, emergency preparedness
• Annual report to respondents
  – Lay summaries of scientific articles
• Calendar of local sites & people
• Study tee shirts (AAH logo)
AAH Study Cohort Maintenance

• Proxies when participants unavailable
  – Temporary or not (admitted to nursing home, incarceration, etc.)
  – Reported deaths verified by vital stats, National Death Index

• Very skilled field staff
  – Interviewers professionals & moved from project to project (e.g., Census work)
  – Extensive supervision, data quality, regular meetings
AAH Study Calendar Photos

Photo credit: K Mickelsen
Measurement Issues in Two Observer Neighborhood Rating Systems

We hypothesized that in addition to individual social & economic circumstances (SES), the neighborhoods people lived in would exert an influence on outcome

… “Place matters”
Baseline Neighborhood Rating Scale

• **Krause**; 5-item assessment

• **10 Interviewers** - general instructions

• **Ratings during baseline household enumeration phase** (before subject recruitment & enrollment)

• **150 ratings repeated** (independent rater)
Krause Rating Method

Undefined / implied neighborhood
<table>
<thead>
<tr>
<th>Krause Scale Items</th>
<th>excellent</th>
<th>good</th>
<th>fair</th>
<th>poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The condition of the houses &amp; buildings</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. The amount of noise from traffic, trains, airplanes, industry, &amp; things like that</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. The quality of the air - amount of pollution, dirt, &amp; fumes in the air</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. Condition of streets &amp; roads</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5. Condition of the yards &amp; sidewalks</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Total score summed, 5-20 points. higher = worse
AAH Psychometric Results: Krause

• Internal consistency alpha = 0.92 (excellent)
• Unidimensional factor; minimal factor loading = 0.80

But a problem surfaced

• Disproportionate ratings of “2” (good)
  – 38% of total scores=10 (all items rated 2)
  – Informal interviewer feedback suggested rating categories were too subjective; & more training needed
But Decent Retest Interrater Reliability

<table>
<thead>
<tr>
<th>Items / Score</th>
<th>Kappa / ICC *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing condition</td>
<td>0.83</td>
</tr>
<tr>
<td>Noise</td>
<td>0.64</td>
</tr>
<tr>
<td>Air quality</td>
<td>0.58</td>
</tr>
<tr>
<td>Streets</td>
<td>0.66</td>
</tr>
<tr>
<td>Yards/sidewalks</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td><strong>0.81</strong></td>
</tr>
</tbody>
</table>

* Intraclass correlation coefficient. 0.75+ is excellent
Decent Discriminant Validity Results

<table>
<thead>
<tr>
<th>Item</th>
<th>Inner city</th>
<th>Suburbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing condition *</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Noise</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Air quality *</td>
<td>2.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Streets *</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Yards/sidewalks *</td>
<td>3.3</td>
<td>2.2</td>
</tr>
<tr>
<td>**Total Score ***</td>
<td><strong>14.9</strong></td>
<td><strong>11.4</strong></td>
</tr>
</tbody>
</table>

* p<0.01 for worse (higher) scores in the inner city
A Problem: Interviewer Effects

Linear regression model (outcome = total score)
- Inner city scores +3.7 points (worse)
- Comparison: experienced U Michigan interviewers
  - New St. Louis interviewers + 0.22 points
  - St. Louis experienced interviewers + 1.8 points

- Individual interviewers still varied by 3.4 points after adjustments
  - (Scale scores 5-20 points / 16 point spread)
AAH Question # 1

Does neighborhood predict incident disability?

• Poor lower-body function plays a crucial role in the disablement process
• Inner-city African Americans have particularly high levels of excess disability & risk for further decline

Research Background

• Balfour & Kaplan (AJE 2002)
  – Persons aged 55 & older who reported residing in neighborhoods with multiple problems were at increased risk of lower-extremity functional loss (odds ratio=3.1) even adjusted by individual measures of social circumstances
  – Neighborhood conditions by self report → possible same-source bias?
The AAH Study

• **Purpose**: To confirm the association shown by Balfour & Kaplan (AJE, 2002) by examining the association between *observed* neighborhood conditions & self-reported *incidence of lower-body functional limitation* (LBFL)
Methods

- **Outcome:** Nagi medical-model LBFL. Self reported
  - difficulties in walking ¼ mile
  - walking up & down 10 steps without rest
  - standing for 2 hours
  - stooping, crouching, or kneeling
  - lifting 10 pounds
- **Subjects with any difficulty or inability to perform the function or task were considered to be limited in that function/task**
- **Sum of functions/tasks (range: 0 - 5)**
Methods: Statistics

- Predict: Development of difficulty/inability to perform 2 or more vs. 0-1 functions/tasks (Balfour & Kaplan)
- logistic regression models & propensity scores for “exposure” to 4-5 fair-poor neighborhood conditions vs. 0-1. And 2-3 vs. 0-1 (dose-response)
- 18 covariates: Social, demographic, health status, behavior

- Various sensitivity analyses
  - Classification of neighborhood condition
  - Classification of lower body functional limitation
  - Method of adjustment for covariates: propensity score
  - Unmeasured binary confounder
Results

- Excluded 290 subjects with 2 or more prevalent LBFL at baseline
- 563 persons with 0-1 LBFL at baseline at risk for 2+ LBFL
- 109 (19%) experienced 2 or more LBFL at 3-year follow-up
## Propensity-adjusted results

<table>
<thead>
<tr>
<th>Multiple conditions rated as fair-poor</th>
<th>Odds ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-5 conditions</td>
<td>3.1</td>
<td>1.6 – 5.9</td>
</tr>
<tr>
<td>2-3 conditions</td>
<td>2.2</td>
<td>1.1 – 4.7</td>
</tr>
<tr>
<td>0-1 conditions</td>
<td>referent</td>
<td></td>
</tr>
</tbody>
</table>
Wave 4 Neighborhood Rating Scale

• 20-item block assessment adapted from the Project on Human Development in Chicago Neighborhoods

Adapted “Chicago” Rating Method

Item Examples: Entire Street

1. **Volume of traffic:**
   - No traffic
   - Light (occasional cars)
   - Moderate
   - Heavy (steady stream of cars)

2. **Condition of street:**
   - Under construction
   - Very poor (many sizable cracks, potholes, broken curbs)
   - Fair
   - Moderately good (no sizable cracks, potholes, broken curbs)
   - Very good
### Item Examples: Block Faces

<table>
<thead>
<tr>
<th>Block faces</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a</strong></td>
</tr>
<tr>
<td>residence</td>
</tr>
</tbody>
</table>

6. **Which of the following are present on the block face?** (yes/no)
   - Abandoned car
   - Empty beer/liquor bottles …

7. **Are there cigarette or cigar butts or discarded cigarette packages on the sidewalk or in gutters?** (yes/no)
AAH Adapted “Chicago”

• Derived 7-item scale (alpha=0.75)

• Extensive photo/rating training protocol developed with St. Louis resident experienced interviewer testing & feedback.
Item 17: Housing condition – Very well kept/good condition – attractive for its type. *Don’t rate cost of the housing, but how it rates for a single-family home of this type.*
“Chicago” Scale Results

Seven Items, one factor

1. Traffic volume
2. Street condition
3. Noise
4. Beer/liquor bottles
5. Cigarettes
6. Garbage, litter
7. Housing condition

Item Factor loadings 0.47-0.80

= items similar to 5-item “Krause”
Results

• No advantage using both block faces
  – Substantial agreement by block face
  – Validity similar for one & two block faces

• Discriminant validity
  – 0-15 point scale scores (one block face)
    6.4 (inner city) vs. 3.5 (suburbs)
But we had interviewer effects

- Linear regression model with total scale as the outcome variable
  - Inner city 3.4 points higher (worse)
  - Interviewer (all St Louis) experience (new vs. experienced) had no effect
  - *Individual interviewers still varied by 4.5 points after adjustment by area* (scores 0-15 / 16-point spread)
Third try is a charm? Wave 10

- Selected 4 best & most experienced interviewers
- 6 hours classroom training
- Test (case photo slides & group discussion)
- 3 hours field training (walk-around as a group)
- Investigator did one field case with each rater
- Investigator reviewed first 51 ratings for quality assurance (returned for questions)
- Forms submitted weekly & reviewed (returns)
Third try is a charm?

- Added observations & data on the duration of rating (minutes), day of week, time, & weather: environmental influences on rating?
- All raters assigned to street segments randomly, then provided map/sequence to rate in clusters (saving time)
- 120 street segments were rated by a randomly selected second rater (intraclass correlation on scale score totals)
## Scale results

<table>
<thead>
<tr>
<th>Scale</th>
<th>Raters ICC*</th>
<th>Discriminant validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Inner City</td>
</tr>
<tr>
<td>Krause (5 item)</td>
<td>0.19</td>
<td>8.6</td>
</tr>
<tr>
<td>AAH (7 items)</td>
<td>0.56</td>
<td>5.2</td>
</tr>
</tbody>
</table>

*ICC 0.75+ is excellent agreement; below 0.30 is poor agreement*
Interviewer results

<table>
<thead>
<tr>
<th>Measures</th>
<th>Interviewer Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
</tr>
<tr>
<td>Time (minutes)</td>
<td>10±7</td>
</tr>
<tr>
<td>Krause</td>
<td>8.0±2.6</td>
</tr>
<tr>
<td>AAH 7-item</td>
<td>4.2±3.0</td>
</tr>
</tbody>
</table>

* Supervisor (somewhat reduced rating assignments)

Linear regression: scores of two interviewers (2, 3) differed significantly from the supervisor (p<0.001) adjusted by area
## Concurrent Validity: Health Outcomes

### Scale scores predict health outcomes? Regression models of the cross sectional data

<table>
<thead>
<tr>
<th>Health Outcome</th>
<th>Association Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower body function (self reported)</td>
<td>Modest association: poorer neighborhood conditions associated with poorer function</td>
</tr>
<tr>
<td>Short physical performance battery</td>
<td>Modest association: poorer neighborhood conditions associated with poorer function</td>
</tr>
<tr>
<td>Peak expiratory flow</td>
<td>Strongest association: poorer neighborhoods associated with poorer lung function</td>
</tr>
</tbody>
</table>

Models adjusted by participant age, gender, area, interviewer
Mismeasurement Effects

- Outcome: participant report self rated health
- Neighborhood conditions assumed to have random additive error (test-retest data)
- Result: Uncorrected & corrected estimates showed lower neighborhood effects associated with lower health
- However, mismeasurement (low retest results) appeared to have driven the association toward the null (reduced the magnitude of the association)
Summary

- Substantial discriminant/construct validity of observer ratings, despite imperfect measures
  - Few items needed for a summary scale
  - Large differences between St Louis areas
- Predictive validity (health outcomes) shows promise, but may be affected by rating error
- Retest reliability was driven down when we increased score variability (we beat the curse of the “2” ratings on Krause)
- Interviewer effects not conquered despite increasing scrutiny to training, reducing number of rater/interviewers
- Few measurement studies published for well-used neighborhood rating measures/systems
Selected Publications of the AAH Study


