# Drug Class Review

# Topical Calcineurin Inhibitors

**Final Report**

**October 2008**

The Agency for Healthcare Research and
Quality has not yet seen or approved this report.

The purpose of this report is to make available information regarding the comparative effectiveness and safety profiles of different drugs within pharmaceutical classes. Reports are not usage guidelines, nor should they be read as an endorsement of, or recommendation for any particular drug, use or approach. Oregon Health & Science University does not recommend or endorse any guideline or recommendation developed by users of these reports.

Nancy J. Lee, PharmD, BCPS
Marian McDonagh, PharmD
Benjamin Chan, MS
Kimberly Peterson, MS
Sujata Thakurta, MPA:HA

Oregon Evidence-based Practice Center
Oregon Health & Science University
Mark Helfand, MD, MPH, Director

Prepared for: Drug Effectiveness Review Project
Principal Investigator, Marian McDonagh, Pharm D



# TABLE OF CONTENTS

**INTRODUCTION ..... 8**
Scope and Key Questions ..... 10
**METHODS ..... 12**
Literature Search ..... 12
Study Selection ..... 12
Data Abstraction ..... 13
Validity Assessment ..... 13
Data Synthesis ..... 14
**RESULTS ..... 17**
Systematic Reviews ..... 19
Key Question 1. ..... 20
Summary ..... 20
Shorter-term treatment (≤12 weeks) ..... 20
Mild to moderate disease ..... 20
Moderate to severe disease ..... 21
Quality of life ..... 21
Active-control trials with topical steroids in moderate to severe disease ..... 21
Maintenance or prevention (24 to 52 weeks) ..... 22
Detailed Assessment ..... 22
Shorter-term treatment (≤12 weeks) ..... 22
Mild to moderate disease ..... 24
Active-control trials with topical steroids in mild to moderate disease ..... 26
Moderate to severe disease ..... 26
Quality of life ..... 28
Active-control trials with topical steroids in moderate to severe disease ..... 29
Maintenance or prevention (24 to 52 weeks) ..... 31
Key Question 2. ..... 32
Summary ..... 32
Detailed Assessment ..... 33
Harms ..... 33
Lymphomas ..... 33
Skin atrophy, telangiectasia, adrenal suppression, or skin striae ..... 34
Withdrawals ..... 34
Application site reactions ..... 35
Herpes simplex virus, molluscum contagiosum, eczema herpeticum, herpes zoster ..... 35
Key Question 3. ..... 36
Summary ..... 36
Detailed Assessment ..... 37
Age ..... 37
Ethnic origin ..... 37
Baseline disease severity ..... 37
Body surface area involved with atopic dermatitis ..... 37
Chronic hand dermatitis ..... 37
Assumptions and Limitations ..... 38
Pooling across populations and stratifying by disease severity ..... 38
Assessment of outcomes ..... 38
Generalizability characteristics ..... 38
**SUMMARY ..... 40**
**REFERENCES ..... 45**

## TABLES

Table 1. Characteristics of tacrolimus and pimecrolimus ..... 9
Table 2. Study inclusion criteria ..... 12
Table 3. Description of included assessment methods ..... 15
Table 4. Summary of results for "investigators' global assessment of response"[a] from a meta-analysis by Ashcroft, et al. ..... 20
Table 5. Head-to-head studies 6 weeks in duration ..... 23
Table 6. Tacrolimus: Vehicle-controlled trials 3 to 12 weeks in duration ..... 23
Table 7. Pimecrolimus: Vehicle-controlled trials 3 to 12 weeks in duration ..... 24
Table 8. Vehicle-controlled trials for indirect comparison of tacrolimus and pimecrolimus (proportion of patients with mild to moderate disease achieving treatment success at the end of 6 weeks) ..... 25
Table 9. Vehicle-controlled trials for indirect comparison of tacrolimus and pimecrolimus (proportion of patients with moderate to severe disease achieving treatment success at the end of 6 weeks) ..... 27
Table 10. Total withdrawal rates for 4 active-control trials ..... 35
Table 11. Number of cases collected from 17 trials (N=7761) that reported serious viral infections ..... 36
Table 12. Definitions of overall strength of evidence ..... 40
Table 13. Summary of the evidence by key question ..... 40

## FIGURES

Figure 1. Results of literature search ..... 18

## APPENDIXES

Appendix A. Search strategy ..... 50
Appendix B. Quality assessment of drug class reviews for the Drug Effectiveness Review Project ..... 52
Appendix C. Excluded studies ..... 57
Appendix D. Glossary ..... 64
Appendix E. Results for sensitivity analysis (Table 8) ..... 71
Appendix F. Assessing the strength of comparative evidence (using the modified GRADE approach) . 72

## EVIDENCE TABLES

**Published in a separate document**

*Suggested Citation for this Report:*

Lee NJ, McDonagh MS, Peterson K, Chan B, Thakurta S. Drug class review: Topical calcineurin inhibitors. Original report. http://www.ohsu.edu/drugeffectiveness/reports/final.cfm

*Funding:*

The funding source, the Center for Evidence-based Policy, is supported by 15 organizations, including 14 state Medicaid programs. These organizations selected the topic and had input into the Key Questions for this review. The content and conclusions of the review are entirely determined by the Evidence-based Practice Center researchers. The authors of this report have no financial interest in any company that makes or distributes the products reviewed in this report.

*Acknowledgments:*

The investigators greatly appreciate the technical assistance provided by Arkady Mak, MD, PhD, our manuscript editor, and Leah Williams-Morris, our publications editor.

We also extend our appreciation to the clinical advisors listed below for their advice and input during our research process:

Jon Hanifin, MD
Oregon Health & Science University

Eric Simpson, MD
Oregon Health & Science University

Jerry Tan, MD
University of Western Ontario

## Purpose and Limitations of Evidence Reports

Systematic reviews or evidence reports are the building blocks of evidence-based practice. An evidence report focuses attention on the strength and limits of evidence from published studies about the effectiveness of a clinical intervention. The development of an evidence report begins with a careful formulation of the problem. The goal is to select questions that are important to patients and clinicians, then to examine how well the scientific literature answers those questions.

An evidence report emphasizes the patient's perspective in the choice of outcome measures. Studies that measure health outcomes (events or conditions that the patient can feel, such as quality of life, functional status, and fractures) are emphasized over studies of intermediate outcomes (such as changes in bone density). An evidence report also emphasizes measures that are easily interpreted in a clinical context. In general, measures of absolute risk or the probability of disease are preferred to measures such as relative risk. The difference in absolute risk between interventions depends on the numbers of events in both groups, such that the difference (absolute risk reduction) is smaller when there are fewer events. In contrast, the difference in relative risk is fairly constant across groups with different baseline risk for the event, such that the difference (relative risk reduction) is similar across these groups. Relative risk reduction is often more impressive than absolute risk reduction. Another measure useful in applying the results of a study is the number needed to treat (or harm). The number needed to treat represents the average number of patients who need to be treated with the intervention of interest in order to achieve 1 additional patient to benefit (that is, experience a positive outcome or avoid a negative outcome) relative to the comparator intervention. The absolute risk reduction is used to calculate the number needed to treat. (For this review, number needed to treat (or harm) were calculated for instances where statistically significant differences were observed between treatment groups, otherwise relative risks were reported).

Evidence reports also consider the quality of the evidence, giving more weight to studies that meet high methodological standards that reduce the likelihood of biased results. In general, for questions about the relative benefits of a drug, the results of well planned and executed randomized controlled trials are regarded as better evidence than results of cohort, case-control, or cross-sectional studies. In turn, these studies are considered better evidence than uncontrolled trials and case series. For questions about tolerability and harms, controlled trials typically provide limited information. For these questions, observational studies may provide important information that is not available from trials. Within this hierarchy, cohort designs are preferred when well conducted and assessing a common outcome. Case-control studies are preferred only when the outcome measure is rare and the study is well conducted.

Evidence reports pay particular attention to the generalizability of results from *efficacy* studies performed in controlled or academic settings to "real world settings." *Efficacy* studies provide the best information about how a drug performs in a controlled setting that allows for better control over potential confounding factors and bias. However, the results of efficacy studies are often not applicable to many, and sometimes most, patients seen in everyday practice. This is because most efficacy studies use strict eligibility criteria that exclude patients based on their age, sex, medication adherence, or severity of illness. For many drug classes, including antipsychotics, unstable or severely impaired patients are often excluded from trials. Efficacy studies also often exclude patients who have comorbid diseases, meaning diseases other than the one under study. Efficacy studies may also use dosing regimens and follow-up protocols that

would be impractical in other practice settings. They often restrict options such as combining therapies or switching drugs that are of value in actual practice. And they often examine the short-term effects of drugs that in practice are used for much longer periods. Finally, efficacy studies tend to use objective measures of effects that do not capture all of the benefits and harms of a drug or do not reflect the outcomes that are most important to patients and their families.

Evidence reports also highlight studies that reflect actual clinical *effectiveness* in unselected patients and community practice settings. *Effectiveness* studies conducted in primary care or office-based settings use less stringent eligibility criteria, assess more health outcomes, and have longer follow-up periods than most efficacy studies. The results of effectiveness studies are more applicable to the "average" patient than results from efficacy studies working with highly selected populations. Examples of effectiveness outcomes include quality of life, hospitalizations, and the ability to work or function in social activities. These outcomes are more important to patients, family, and care providers than surrogate or intermediate measures such as scores on psychometric scales.

Efficacy and effectiveness studies overlap. For example, a study might use very narrow inclusion criteria like an efficacy study, but like an effectiveness study might examine flexible dosing regimens, have a long follow-up period, and measure quality of life and functional outcomes. For this report we sought evidence about outcomes that are important to patients and would normally be considered appropriate for an effectiveness study. However, many of the studies that reported these outcomes were short-term and used strict inclusion criteria to select eligible patients. It was neither possible nor desirable to exclude evidence from these studies. Labeling each study as either an efficacy or an effectiveness study, while convenient, is of limited value; it is more useful to consider whether the patient population, interventions, time frame, and outcomes are relevant to one's practice or to a particular patient.

Studies across the continuum from efficacy to effectiveness can be useful in comparing the clinical value of different drugs. Effectiveness studies are more applicable to practice, but efficacy studies are a useful scientific standard to determine whether the characteristics of different drugs are related to their effects on disease. An evidence report reviews the efficacy data thoroughly to ensure that decision makers can assess the scope, quality, and relevance of available data. This thoroughness is not intended to obscure the fact that efficacy data, no matter how much of it there is, may have limited applicability to practice. Clinicians can judge the relevance of study results to their practice and should note where there are gaps in the available scientific information.

Unfortunately, for many drugs there are few or no effectiveness studies and many efficacy studies. Consequently, clinicians must make decisions about treatment for many patients who would not have been included in controlled trials and for whom the effectiveness and tolerability of different drugs are uncertain. An evidence report indicates whether or not there is evidence that drugs differ in their effects in various subgroups of patients, but it does not attempt to set a standard for how results of controlled trials should be applied to patients who would not have been eligible for them. With or without an evidence report, these decisions must be informed by clinical judgment.

In the context of developing recommendations for practice, evidence reports are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. By themselves, they do not say what to do. Judgment, reasoning, and applying one's values under conditions of uncertainty must also play a role in decision making. Users of an evidence report must also keep

in mind that *not proven* does not mean *proven not*; that is, if the evidence supporting an assertion is insufficient, it does not mean the assertion is not true. The quality of the evidence on effectiveness is a key component, but not the only component, in making decisions about clinical policies. Additional criteria include acceptability to physicians or patients, the potential for unrecognized harms, the applicability of the evidence to practice, and consideration of equity and justice.

# INTRODUCTION

Atopic dermatitis, also referred to as atopic eczema, is a highly pruritic, chronic, and relapsing inflammatory disease of the skin. The natural history of the disease is not fully understood, and popular notions about the etiology of atopic dermatitis such as "hygiene theory" and "atopic march" are continuing to be reassessed.[1] In general, atopic dermatitis primarily affects infants, children, and adolescents with a 15% to 30% prevalence, compared with 2% to 10% prevalence seen in adults.[2, 3] Approximately 45% of all cases occur during the first year of life and 85% of cases occur before 5 years of age.[4] Young children typically exhibit more severe and persistent disease than older patients, although new-onset atopic dermatitis in adults is possible. Furthermore, periods of remission usually occur more frequently as age increases. It is estimated that 40% to 70% of children older than 5 years may experience spontaneous resolution of their disease.[2, 5] Of these cases, however, more than 50% of patients can relapse back to active atopic dermatitis, though the severity may not be as intense.[4]

Atopic dermatitis is associated with significant weakening of the skin barrier, allowing for increased susceptibility to water loss, to allergens, and to infectious pathogens. Several genes that encode proinflammatory cytokines (responsible for increased IgE and IgG activity via T-lymphocytes) have also been identified and linked to the complex pathogenesis of atopic dermatitis.[4, 6] Diagnosis of atopic dermatitis is based on a constellation of symptoms. The essential feature is pruritus, which provokes a vicious itch-scratch-rash cycle. Patient and family history of atopy and recurrent eczematous lesions are additional features involved in diagnosis.[6] Currently, however, there is no standard method for diagnosis and various criteria are used. The most commonly cited criteria is the Hanifin and Rajka criteria; however strong arguments in favor of using the Sampson or the Williams criteria in children have also been made.[7]

There is no known cure for atopic dermatitis and no optimal regimen for long-term maintenance of the disease.[8] Treatment of atopic dermatitis usually involves a multipronged approach of reducing exposure to exacerbating factors, maintaining skin hydration with emollients, alleviating symptoms such as pruritus, and controlling active disease with topical anti-inflammatory agents.[6] Intensity of treatment with or without a topical anti-inflammatory agent depends on the severity of the disease. Of the topical agents, topical steroids are generally considered the mainstay of treatment. Until recently, the use of low- to mid-potency topical steroids has been recommended for maintenance therapy, whereas high-potency agents have been reserved for significant flares.[6] Currently, several different treatment regimens using mid- to high-potency topical steroids dosed less frequently are being implemented in clinical practice.[1, 8, 9] Despite the shift in topical steroid use, concerns about side effects associated with long-term topical steroid exposure continue to persist among patients and practitioners. Hence, treatments with alternate nonsteroid based agents are being sought.

In December 2000 and 2001, two topical calcineurin inhibitors were approved for use in patients with atopic dermatitis in the United States and Canada. (See Table 1 for mechanism of action). Since the approval of these agents, several case reports of malignancies (skin and lymphoma) have been reported to the United States Food and Drug Administration, causing a black box warning to be placed in each product's labeling. Several pharmacokinetic analyses, commentaries, and editorials have been published refuting the addition of the black box warning. In light of these findings, this comparative effectiveness review of 2 topical calcineurin inhibitors was commissioned to identify whether additional good-quality studies on safety have been

published and to determine whether differences in efficacy and effectiveness exist between the 2 topical agents.

### Table 1. Characteristics of tacrolimus and pimecrolimus

| | | |
|---|---|---|
| **Scientific name** | Tacrolimus | Pimecrolimus |
| **Brand** | Protopic® | Elidel® |
| **Chemical structure** | Macrolide | Ascomycin derivative |
| **Manufacturer** | Astellas Pharma | Novartis |
| **Approval date** | December 8, 2000 | December 13, 2001 |
| **Country** | US, Canada | US, Canada |
| **Dose** | 0.03%, 0.1% | 1% |
| **How supplied** | Ointment | Cream |
| **FDA Indication** | Children (2 to 15 years): 0.03%<br>Adults: 0.03%, 0.1%<br><br>Indicated as *second-line therapy* for the short-term and noncontinuous chronic treatment of *moderate to severe* atopic dermatitis in non-immunocompromised adults and children who have failed to respond adequately to other topical prescription treatments for atopic dermatitis, or when those treatments are not advisable.<br><br>Not indicated for children younger than 2 years of age. | Children (2 to 15 years) and Adults: 1%<br><br>Indicated as *second-line therapy* for the short-term and noncontinuous chronic treatment of *mild to moderate* atopic dermatitis in unimmunocompromised adults and children 2 years of age and older, who have failed to respond adequately to other topical prescription treatments, or when those treatments are not advisable.<br><br>Not indicated for use in children less than 2 years of age. |
| **Black box warning** | Although a causal relationship has not been established, rare cases of malignancy (e.g., skin and lymphoma) have been reported in patients treated with topical calcineurin inhibitors, including pimecrolimus and tacrolimus. Therefore, continuous long-term use of topical calcineurin inhibitors in any age group should be avoided, and application limited to areas of involvement with atopic dermatitis. | |
| **Precautions** | Should be avoided on malignant or premalignant skin conditions. Malignant or premalignant skin conditions, such as cutaneous T-cell lymphoma (CTCL), can present as dermatitis.<br><br>Should not be used in patients with Netherton syndrome or other skin diseases where there is the potential for increased systemic absorption of pimecrolimus or tacrolimus. The safety of pimecrolimus or tacrolimus has not been established in patients with generalized erythroderma. | |
| **Contraindications** | Contraindicated in individuals with a history of hypersensitivity to tacrolimus or pimecrolimus or any of the components of the cream or ointment. | |
| **Mechanism of action** | The mechanism of action of tacrolimus in atopic dermatitis is not known. Tacrolimus has been shown to inhibit T-lymphocyte activation by first binding to intracellular protein macrophilin-12 (also known as FKBP-12). A complex of tacrolimus-FKBP-12, calcium, calmodulin, and | The mechanism of action of pimecrolimus in atopic dermatitis is not known. Pimecrolimus has been shown to bind with high affinity to macrophilin-12 ( also known as FKBP-12) and inhibit calcineurin. As a consequence, it inhibits T cell activation by blocking transcription of early cytokines. In particular, nanomolar concentrations of |

| | | |
|---|---|---|
| | calcineurin is then formed and the phosphatase activity of calcineurin is inhibited. This effect has been shown to prevent the dephosphorylation and translocation of nuclear factor of activated T-cells (NF-AT), a nuclear component thought to initiate gene transcription for the formation of lymphokines (such as IL-2 and gamma interferon). Tacrolimus also inhibits transcription of genes encoding IL-3, IL-4, IL-5, GM-CSF, and TNF-α, all of which are involved in the early stages of T-cell activation. Additionally, tacrolimus has been shown to inhibit the release of pre-formed mediators from skin mast cells and basophils, and to down regulate the expression of FcεRI on Langerhans cells. | pimecrolimus inhibit synthesis of IL-2 and interferon gamma (Th1-type) and IL-4 and IL-10 (Th2-type) cytokine synthesis in human T cells. In addition, pimecrolimus prevents release of inflammatory cytokines and mediators from mast cells in vitro after stimulation by antigen/IgE. |

Abbreviations: GM-CSF, granulocyte-macrophage colony stimulating factor; IL, interleukin; TNF, tumor necrosis factor.

## Scope and Key Questions

The purpose of this review is to compare the effectiveness and harms of topical calcineurin inhibitors in persons with atopic dermatitis or eczema. The key questions for this review were developed with input from experts in the field of dermatology. The Oregon Evidence-based Practice Center wrote preliminary key questions identifying the populations, interventions, and outcomes of interest and, based on these, the eligibility criteria for studies. The key questions were reviewed and revised by representatives of organizations participating in the Drug Effectiveness Review Project. The participating organizations of the Drug Effectiveness Review Project were responsible for ensuring that the scope of the review reflected the populations, drugs, and outcome measures of interest to clinicians and their patients. The participating organizations approved the following key questions to guide this review:

### Key Questions

1. For adults and children with stable atopic dermatitis or eczema, do pimecrolimus and tacrolimus differ in effectiveness when compared to each other and when compared to topical corticosteroids:
    a. depending on location of application (for example, head and neck, flexures, hands, feet, intertriginous regions)?
    b. depending on body surface area involved?
    c. depending on treatment duration?

2. For adults and children with stable atopic dermatitis or eczema, do pimecrolimus or tacrolimus differ in safety or adverse events when compared to each other and when compared to topical corticosteroids:
    a. depending on location of application (for example, head and neck, flexures, hands, feet, intertriginous regions)?

   b. depending on body surface area involved?
   c. depending on treatment duration?

3. Are there other subgroups of patients based on demographics (for example, age, racial groups, gender) and comorbidities (for example, immunodeficiencies) for which either pimecrolimus or tacrolimus is more effective or associated with fewer adverse events?

# METHODS

## Literature Search

To identify relevant citations, we searched Ovid MEDLINE® (1950 to November week 2, 2007), the Cochrane Database of Systematic Reviews® (4th quarter 2007), and the Cochrane Central Register of Controlled Trials® (4th quarter 2007) using terms for included drugs, indications, and study designs. (See Appendix A for complete search strategies). We attempted to identify additional studies through hand searches of reference lists of included studies and reviews. In addition, we searched the United States Food and Drug Administration's Center for Drug Evaluation and Research web site for medical and statistical reviews of individual drug products (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/). Finally, we requested dossiers of published and unpublished information from relevant pharmaceutical companies. All received dossiers were screened for studies or data not found through other searches. All citations were imported into an electronic database (Endnote® version 11.0).

## Study Selection

All citations were reviewed for inclusion using the criteria shown in Table 2. One investigator reviewed titles and abstracts of citations while another investigator double-checked the selected references. Full-text articles were retrieved and again were assessed for inclusion by two reviewers; disagreements were resolved by consensus. Results published in abstract form (for example, as a conference proceeding) were not included because these typically do not provide sufficient detail to perform adequate quality assessment. Case reports, case series, and single-arm extension studies also were excluded.

### Table 2. Study inclusion criteria

| |
|---|
| Populations |
| • Adults and children (all ages, including infants) with stable atopic dermatitis or eczema |
| Interventions |
| • Pimecrolimus (Elidel®)<br>• Tacrolimus (Protopic®)<br>Indirect comparators<br>• Placebo<br>• Topical corticosteroids |
| Efficacy of effectiveness outcomes |
| • Frequency of rebound flare-ups<br>• Reduction in symptom severity (for example, sleep loss, pruritus)<br>• Duration of effectiveness (for example, time to next flare-up)<br>• Quality of life<br>• Treatment failure (for example, use of alternative treatments) |

| Harms-related outcomes |
|---|
| • Overall adverse events reported<br>• Withdrawals<br>• Withdrawals due to adverse events<br>• General adverse events (for example, burning, stinging)<br>• Major adverse events (for example, cancers, infections, glaucoma, sensitivity to temperature changes, cutaneous atrophy) |
| **Study designs** |
| • For effectiveness: or randomized controlled trial with duration of ≥3 weeks or good-quality systematic review<br>For harms: randomized controlled trials with duration of ≥3 weeks, good-quality systematic review, observational study (cohort including database studies with comparison group, case-control, before-after studies) with duration of ≥3 weeks. |

## Data Abstraction

The following data were abstracted by one reviewer and reviewed by a second: study design, setting and population characteristics (including sex, age, ethnicity, and diagnosis); eligibility and exclusion criteria; interventions (dose and duration); comparisons; numbers screened, eligible, enrolled, and lost to follow-up; method of outcome ascertainment; and results for each outcome. We recorded intention-to-treat results when reported.

For included systematic reviews, we abstracted the searched databases, study eligibility criteria, number of studies and patients represented, characteristics of included studies, data synthesis methods, and main efficacy and safety results.

## Validity Assessment

We assessed the internal validity (quality) of trials on the basis of the predefined criteria listed in Appendix B. These criteria are based on the United States Preventive Services Task Force and the National Health Service Centre for Reviews and Dissemination (United Kingdom) criteria.[10, 11] We rated the internal validity of each trial on the basis of the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; adequate reporting of dropouts, attrition, crossover, adherence, and contamination; loss to follow-up; and the use of intention-to-treat analysis. We considered methods to meet criteria for intention-to-treat analysis if outcomes for at least 95% of participants were analyzed according to the group to which they were originally assigned. We considered total attrition of ≥20% in any of the treatment arms to be excessive.

Trials that had fatal flaws were rated poor-quality. Trials that met *all* criteria were rated good-quality and the remainder rated fair-quality. As the fair-quality category is broad, studies with this rating vary in their strengths and weaknesses. The results of some fair-quality studies are *likely* to be valid, while others are only *probably* valid. A poor-quality trial is not valid; the results are at least as likely to reflect flaws in the study design as the true difference between the compared drugs. A fatal flaw is reflected by failing to meet *combinations* of items of the quality assessment checklist that work together to suggest a potential for bias.

We assessed the quality of systematic reviews using predefined criteria developed by Oxman and Guyatt (see Appendix B). These included adequacy of literature search and study

selection methods, methods of assessing validity of included trials, methods used to combine studies, and validity of conclusions.

## Data Synthesis

We constructed evidence tables showing the study characteristics, quality ratings, and results for all included studies. We reviewed studies using a hierarchy of evidence approach, where the best evidence was the focus of our synthesis for each question, population, intervention, and outcome addressed. Studies that evaluated one topical calcineurin inhibitor against another provided direct evidence of comparative effectiveness and adverse event rates. These direct comparisons were preferred over indirect comparisons. When available, these data were the primary focus. Similarly, effectiveness and long-term harms-related outcomes were preferred to efficacy and short-term tolerability outcomes.

In theory, trials that compared topical calcineurin inhibitors to other drug classes or placebo could provide evidence about comparative effectiveness. But such indirect comparisons can be difficult to interpret for a number of reasons, including heterogeneity between trial populations, interventions, and assessment of outcomes. Data from indirect comparisons were used to support direct comparisons, where they existed, and were also used as the primary comparison where no direct comparisons existed. Thus, indirect comparisons should be interpreted with caution.

Meta-analyses in this review were conducted using random effects model for outcomes for which a sufficient number of studies existed and for studies that were homogeneous enough that combining their results could be justified.[12] In order to determine whether meta-analysis could be meaningfully performed, we considered the study quality and heterogeneity in design, patient population, interventions, and outcomes. An adjusted indirect comparison was performed for the outcome of resolution of disease assessed by patients by combining the results of the meta-analysis comparing tacrolimus versus vehicle with the meta-analysis comparing pimecrolimus versus vehicle. The variance of the estimate of effect was estimated as the sum of the variances of the two meta-analyses being pooled.[13] Weighted mean differences between drug and control were also calculated for outcomes (change in pruritus, change in EASI score, etc). Incidence rates between drug and control were pooled for withdrawals. The Q statistic and the $I^2$ statistic (the proportion of variation in study estimates due to heterogeneity) were also calculated to assess heterogeneity between the effects from the studies.[14, 15] Analyses were conducted using "R statistical environment" and StatsDirect (CamCode, Altrincham UK) software.

We included studies with adults and children (all ages, including infants) with atopic dermatitis. Publications that pooled more than 1 trial and also provided individual trial results were included, and the data from these trials were used for our meta-analyses. We stratified data by disease severity (mild-moderate compared with moderate-severe), by treatment duration (≤12 weeks compared with >24 weeks), and by tacrolimus strength (0.03% compared with 0.1%). Our decision to stratify by tacrolimus strength was based on our indirect meta-analysis of 5 tacrolimus studies which included tacrolimus 0.03% ointment and 0.1% ointment arms.[16-19] Efficacy and effectiveness outcomes that are reported in this review are Investigator Global Assessment-Atopic Dermatitis (IGA) score ≤1, Physician Global Evaluation (PGE) 90% to 100% improvement, patient assessment of pruritus, patient assessment of overall disease control, percent of patients without flares, time to first flare, percent of patients not using topical steroid rescue, and quality of life. In instances where Eczema Area and Severity Index (EASI) scores were reported similarly enough across trials for comparisons to be made, these data were

reported. In this review, we defined treatment success as either achievement of IGA score ≤1 or achievement of PGE of 90% to 100% improvement in disease from baseline. And if IGA scores were reported as percent achieving a score ≤1, we combined this data with the percent of patients reporting improvements in PGE of 90 to 100%. Table 3 provides a brief description of IGA, PGE, and EASI scoring methods.

### Table 3. Description of included assessment methods

<table>
<tr><th>Assessment methods</th><th>Validated?</th><th>Description</th></tr>
<tr><td>Investigator Global Assessment-Atopic Dermatitis (IGA)</td><td>Partially[20]</td><td>Static 6-point scale based on assessment of erythema and infiltration/papulation from 0 (clear) to 5 (very severe disease)<br><br>In most trials, scores ≤1 were generally classified as “treatment success,” whereas scores >1 were considered “treatment failure.”<br><br><table><tr><td>0-clear</td><td>No inflammatory signs of disease</td></tr><tr><td>1-almost clear</td><td>Just perceptible erythema and infiltration/papulation</td></tr><tr><td>2-mild disease</td><td>Mild erythema and infiltration/papulation</td></tr><tr><td>3-moderate disease</td><td>Moderate erythema and infiltration/papulation</td></tr><tr><td>4-severe disease</td><td>Severe erythema and infiltration/papulation</td></tr><tr><td>5-very severe disease</td><td>Severe erythema and infiltration/papulation with oozing/crusting</td></tr></table></td></tr>
<tr><td>Physician Global Evaluation (PGE)</td><td>Unknown</td><td>Change in clinical status scored as percent improvement of lesions identified for treatment at baseline.<br><br>Typically, “success” was defined as ≥ 90% improvement of the monitored lesions.<br><table><tr><th>Improvement</th><th></th></tr><tr><td>100%</td><td>Cleared</td></tr><tr><td>90% to 99%</td><td>Excellent improvement</td></tr><tr><td>75% to 89%</td><td>Marked improvement</td></tr><tr><td>50% to 74%</td><td>Moderate improvement</td></tr><tr><td>30% to 49%</td><td>Slight improvement</td></tr><tr><td>0% to 29%</td><td>No improvement</td></tr><tr><td><0</td><td>Worse</td></tr></table></td></tr>
</table>

<table>
<tr><th>Assessment methods</th><th>Validated?</th><th>Description</th></tr>
<tr><td>Eczema Area and Severity Index (EASI)[21]</td><td>Yes[20]</td><td>4-point (0, none; 1, mild; 2, moderate; 3, severe) scale assessing erythema, infiltration/papulation, excoriation, and lichenification separately on the head/neck, trunk, upper limbs, and lower limbs.<br><br>EASI assigns proportionate values to each of the 4 body regions (roughly based on the “rule of nines”). The overall score ranges from 0 (no disease) to 72 (all signs of disease rated severe and present on 100% of body surface area).<br><br>
<table>
<tr><th>Body regions</th><th>Scoring formula[a]</th></tr>
<tr><td>Upper limbs</td><td>(Eryth+Infil+Excor+Lich) x area involved x 0.2</td></tr>
<tr><td>Lower limbs</td><td>(Eryth+Infil+Excor+Lich) x area involved x 0.4</td></tr>
<tr><td>Trunk</td><td>(Eryth+Infil+Excor+Lich) x area involved x 0.3</td></tr>
<tr><td>Head/neck</td><td>(Eryth+Infil+Excor+Lich) x area involved x 0.1</td></tr>
</table>
[a] area involved within each body region was estimated as the percentage of the total area of that region.</td></tr>
</table>

## Peer Review and Public Comment

Original Drug Effectiveness Review Project reports are independently reviewed and commented upon by three to five peer reviewers. Peer reviewers are identified through a number of sources, including but not limited to members of professional societies, acknowledged experts in a particular field, authors figuring prominently in the published literature, and persons recommended by the Drug Effectiveness Review Project participating organizations. A list of peer reviewers for Drug Effectiveness Review Project reports is available on the Drug Effectiveness Review Project website (www.ohsu.edu/drugeffectiveness).

The Drug Effectiveness Review Project process allows for a two-week public comment period prior to finalization of the report. Draft reports are posted on the Drug Effectiveness Review Project website and interested individuals or organizations can submit comments. Comments received from peer reviewers are considered and revisions made accordingly. Public comments are discussed with the Drug Effectiveness Review Project participating organizations and then a determination is made as to what revisions are appropriate.

# RESULTS

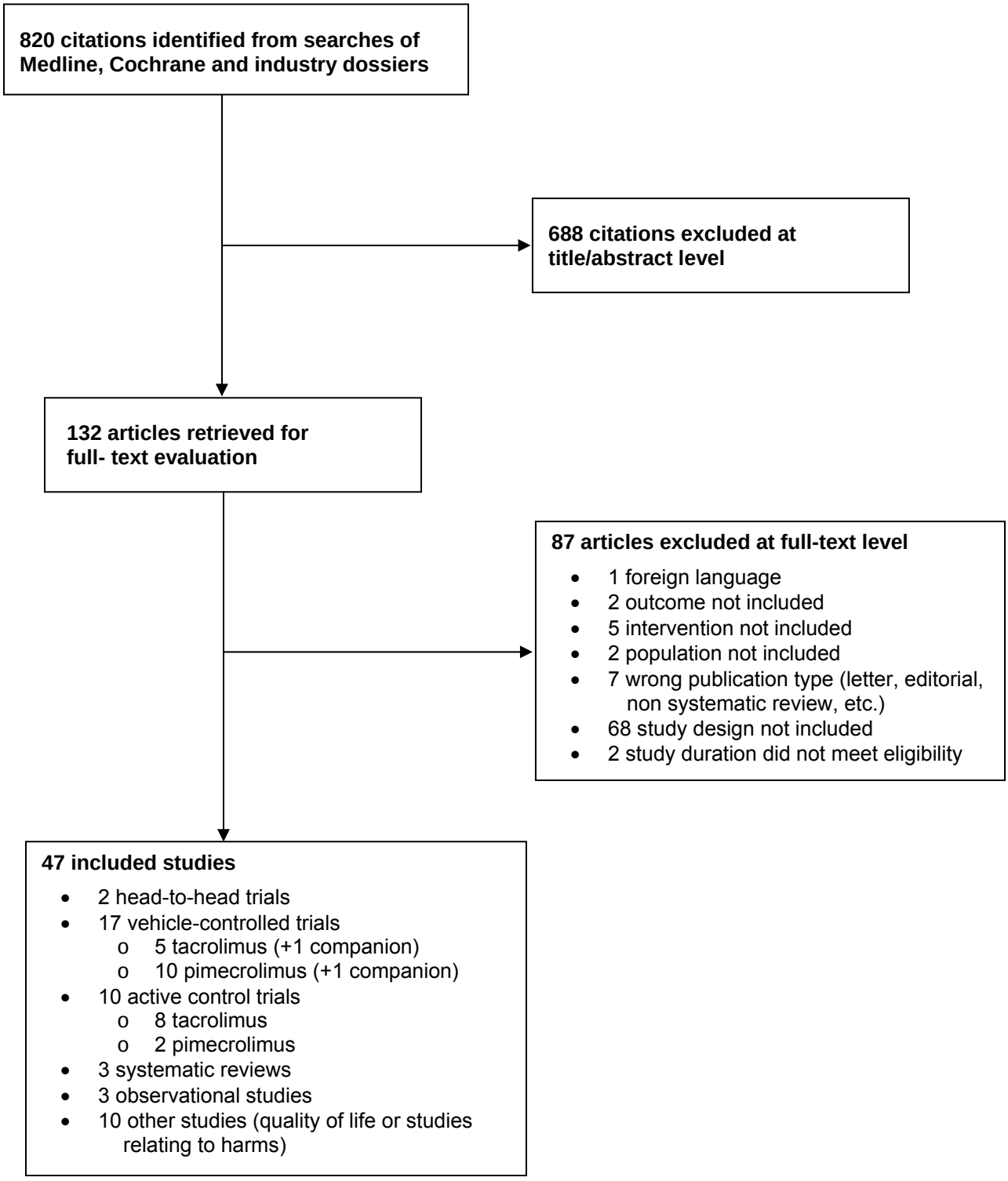
We identified 820 citations by literature searches and 45 studies were included. Ten relevant trials were also identified from the United States Food and Drug Administration (US FDA) Medical and Statistical Reviews. Of these trials, 5 have been published in peer reviewed journals and are included. The remaining trials could not be found as separate publications, but results from 4 trials (study #35, #36, #305, and #307) were discovered in 2 pooled analyses, which we included.[16, 22] Data from the 2 pooled analyses were verified with information found in the FDA Medical and Statistical Reviews.

Figure 1 shows a breakdown of included studies. Overall, the comparative evidence base is largely from short-term studies that were no longer than 12 weeks in duration: 17 vehicle-controlled trials, 10 active-control studies, and 2 head-to-head publications. Only 5 vehicle-controlled trials studied pimecrolimus over 24 to 52 weeks and assessed longer term outcomes.

More than two-thirds of the trials were conducted in children (2 to 15 years) and infants (3 to 23 months). Of tacrolimus vehicle-controlled trials, all but 1 trial was conducted in patients with moderate to severe disease[16, 17, 23, 24] whereas only 2 pimecrolimus vehicle-controlled trials were conducted in patients with moderate to severe disease.[25, 26] At baseline atopic dermatitis affected 10% to 40% of total body surface area with more than 70% of affected surface area located on head and neck region. Fifty to seventy percent of those enrolled were white and female; 20% to 30% were black. Patients with concomitant infections or significant comorbid conditions (such as Netherton syndrome) were excluded from trials. Patients were mostly recruited from dermatology or allergy clinics and were likely managed by specialists (for example, dermatologists).

With the exception of 3 active-control studies, all other trials were rated fair-quality. The 3 active-control trials[27-29] were rated poor-quality based on a combination of factors: inadequate randomization, unclear allocation concealment, and unclear or inadequate blinding (Evidence Table 9).

## Figure 1. Results of literature search



## Systematic Reviews

Three systematic reviews[30-32] were included. The 3 reviews included evidence on tacrolimus and pimecrolimus and outcomes were fairly similar to the scope of our review. Of the 3 reviews only 1 was the most recently published (last search date December 2004). Results from this systematic review and meta-analysis are summarized below.[33]

Twenty-five studies which included randomized trials, abstracts, and non-English publications were included in the systematic review and meta-analysis by Ashcroft, et al. Of these, 11 trials assessed pimecrolimus 1% cream (8 vehicle-controlled, 2 active-control, 1 head-to-head) and 14 trials assessed tacrolimus 0.03% or 0.1% ointment (7 vehicle-controlled, 7 active-control) in patients with varying degrees of atopic dermatitis severity. The primary outcome was a combination of 2 similar endpoints: 1) the proportion of patients with clear or almost clear resolution of disease as assessed by investigators per IGA score ≤1 for pimecrolimus trials and 2) the proportion of patients who achieved at least a 90% improvement in their lesions from baseline per PGE 90% to 100% for tacrolimus trials. Ashcroft and colleagues refer to the combined outcome as "investigators' global assessment of response." Secondary efficacy outcomes included: patients' assessment of disease, proportion of patients with flares, and improvement in quality of life. Adjusted indirect meta-analyses comparing tacrolimus ointment with pimecrolimus cream was not performed but analyses comparing tacrolimus or pimecrolimus with vehicle were conducted. Studies were stratified by treatment duration. Active-control trials with topical steroids as the comparator were grouped by the relative topical steroid potency.

Table 4 provides a summary of the results for the primary outcome. Compared with vehicle, treatment with tacrolimus (0.03% or 0.1%) ointment or pimecrolimus 1% cream was superior. One head-to-head study, however, found no significant difference between tacrolimus 0.03% ointment and pimecrolimus 1% cream at 6 weeks. Compared with relatively more potent topical steroids (betamethasone valerate, hydrocortisone butyrate), tacrolimus 0.03% ointment and pimecrolimus 1% cream was less effective. When tacrolimus 0.1% ointment was compared with hydrocortisone butyrate with or without hydrocortisone acetate, tacrolimus was more effective.