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

TABLES

FIGURES

APPENDIXES

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% Adults: 0.03%, 0.1% 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. Not indicated for children younger than 2 years of age.	Children (2 to 15 years) and Adults: 1% 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. 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. 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®) • Tacrolimus (Protopic®) Indirect comparators • Placebo • Topical corticosteroids
Efficacy of effectiveness outcomes
• Frequency of rebound flare-ups • Reduction in symptom severity (for example, sleep loss, pruritus) • Duration of effectiveness (for example, time to next flare-up) • Quality of life • Treatment failure (for example, use of alternative treatments)

Harms-related outcomes
• Overall adverse events reported • Withdrawals • Withdrawals due to adverse events • General adverse events (for example, burning, stinging) • 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 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)

In most trials, scores ≤1 were generally classified as “treatment success,” whereas scores >1 were considered “treatment failure.”

<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.

Typically, “success” was defined as ≥ 90% improvement of the monitored lesions.
<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.

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).

<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

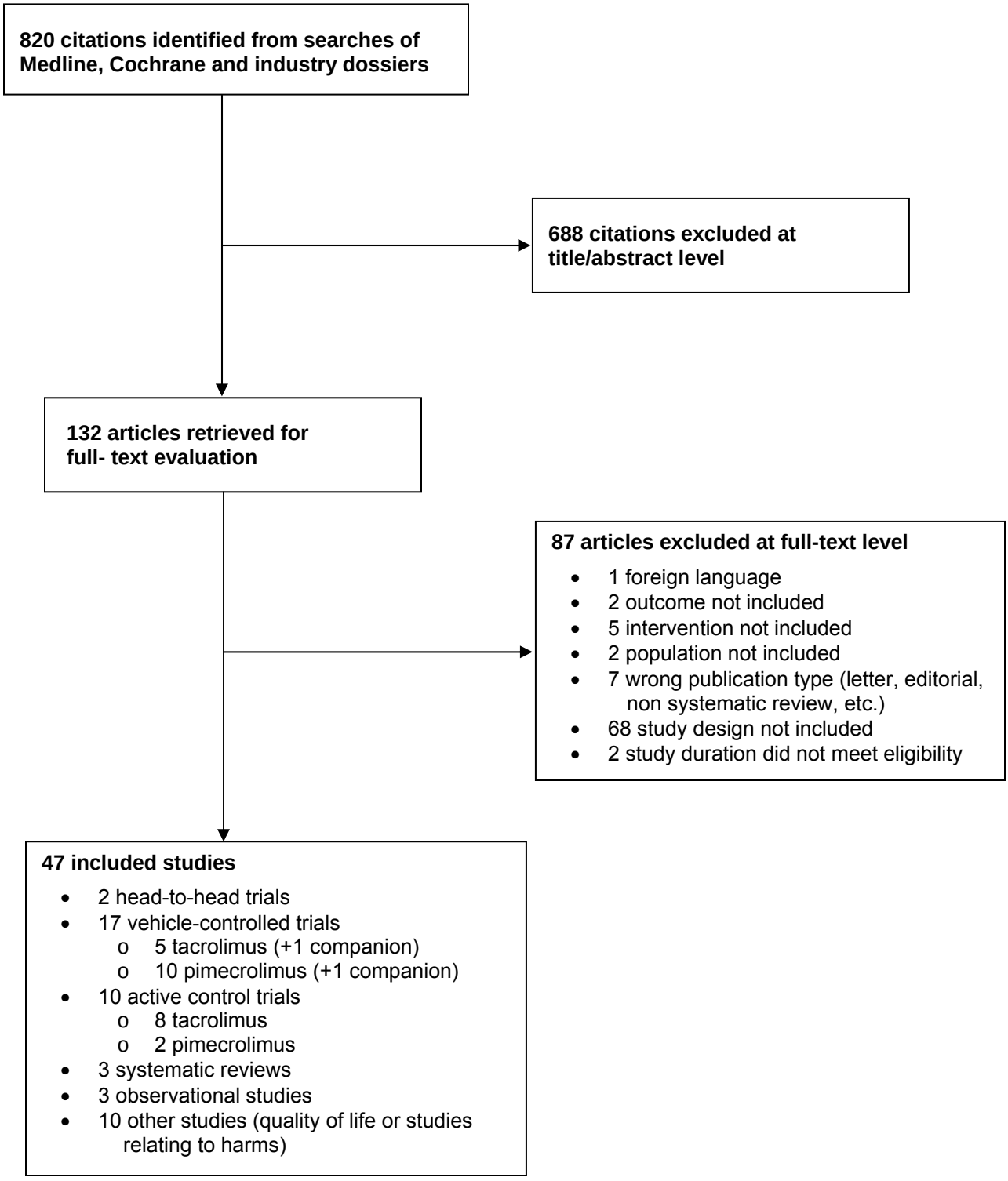
We identified 820 citations by literature searches and 45 studies were included. Ten relevant trials were also identified from the United States Food and Drug Administration (US FDA) Medical and Statistical Reviews. Of these trials, 5 have been published in peer reviewed journals and are included. The 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.

Table 4. Summary of results for "investigators' global assessment of response"[a] from a meta-analysis by Ashcroft, et al.

	Number of included studies	Relative risk[b] (95% CI)
At 3 weeks		
Pimecrolimus compared with vehicle	5	2.72 (1.84 to 4.03)
Pimecrolimus compared with BMV 0.1%	1	0.22 (0.09 to 0.54)
Tacrolimus 0.03% compared with vehicle	1	2.13 (1.24 to 3.68)
Tacrolimus 0.1% compared with vehicle	1	1.57 (0.88 to 2.81)
Tacrolimus 0.03% compared with tacrolimus 0.1%	3	0.89 (0.67 to 1.19)
Tacrolimus 0.03% compared with HCA 1%	2	2.56 (1.95 to 3.36)
Tacrolimus 0.1% compared with HCA 1%	1	3.05 (2.12 to 4.40)
Tacrolimus 0.03% compared with HB 0.1%	1	0.73 (0.58 to 0.93)
Tacrolimus 0.1% compared with HB 0.1%	1	0.95 (0.78 to 1.17)
At 6 weeks		
Pimecrolimus compared with vehicle	3	2.03 (1.50 to 2.74)
Pimecrolimus compared with tacrolimus 0.03%	1	0.71 (0.45 to 1.12)
At 12 weeks		
Tacrolimus 0.03% compared with vehicle	3	4.50 (2.91 to 6.96)
Tacrolimus 0.1% compared with vehicle	3	5.62 (3.67 to 8.61)
Tacrolimus 0.03% compared with tacrolimus 0.1%	3	0.80 (0.65 to 0.99)
Tacrolimus 0.1% compared with HB 0.1%+HCA 1%	1	1.67 (1.41 to 1.98)

[a] For pimecrolimus trials, "investigators' global assessment of response" was determined by using IGA score ≤1 (clear or almost clear resolution of disease); for tacrolimus trials, "investigators' global assessment of response" was determined as >90% improvement from baseline.
[b] Referred to as "rate ratio" in the publication.
Abbreviations: BMV, betamethasone valerate; HB, hydrocortisone butyrate; HCA, hydrocortisone acetate.

Key Question 1.

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

Summary

Shorter-term treatment (≤12 weeks)

Mild to moderate disease

Tacrolimus 0.03% ointment was as effective as pimecrolimus 1% cream in treating atopic dermatitis in patients with mild to moderate disease in 2 head-to-head studies (pooled relative risk, 1.19, 95% CI 0.98 to 1.45) and in an indirect meta-analysis of 4 vehicle-controlled trials

(pooled relative risk 0.97, 95% CI 0.63 to 1.48). Improvements in pruritus were also not significantly different between tacrolimus 0.03% ointment and pimecrolimus 1% cream in the 2 head-to-head studies. None of the studies that included mild to moderate disease severity reported patients' assessment of overall disease control and none stratified efficacy outcomes depending on affected body surface area. Evidence evaluating treatment effect in the head/neck area was found but is limited. Only 1 tacrolimus trial and 1 pooled pimecrolimus study reported mean EASI score improvement which suggests that pimecrolimus 1% cream may be slightly more, or as effective as tacrolimus 0.03% ointment in the head/neck region. We did not find any studies that investigated higher strength tacrolimus 0.1% ointment with pimecrolimus 1% cream in patients with mild to moderate disease. No active-control studies comparing tacrolimus (0.03%, 0.1%) ointment with topical steroids in this population were identified.

Moderate to severe disease

There is insufficient head-to-head evidence comparing lower strength tacrolimus 0.03% ointment with pimecrolimus 1% cream in patients with moderate to severe disease. Indirect comparison of tacrolimus 0.03% ointment and pimecrolimus 1% cream showed no statistically significant difference in achieving treatment success at the end of 3 to 12 weeks (pooled relative risk 0.89, 95% CI 0.38 to 2.07). There was also little difference in pruritus score (pooled weighted mean difference 0.86, 95% CI -0.69 to 2.41) or patient assessment of overall disease control (pooled relative risk 0.98, 95% CI 0.56 to 1.73) between the topical agents.

Direct and indirect evidence reported conflicting results when higher strength tacrolimus 0.1% ointment was compared with pimecrolimus 1% cream in patients with moderate to severe disease. Tacrolimus 0.1% ointment was shown to be more effective in achieving treatment success than pimecrolimus 1% cream (relative risk 1.83, 95% CI 1.13 to 2.96) and maybe slightly better at relieving pruritus (-3.7 cm compared with -2.0 cm, $P \leq 0.01$) at 6 weeks in a small head-to-head study. In contrast, indirect comparison of 3 tacrolimus and 1 pimecrolimus trial revealed no statistically significant differences in treatment success (pooled relative risk 1.17, 95% CI 0.43 to 3.20), change in pruritus score (pooled weighted mean difference 0.74, 95% CI -0.83 to 2.31), or differences in patients' assessment of overall disease control (pooled relative risk 1.07, 95% CI 0.53 to 2.13).

Results from indirect meta-analyses for both mild to moderate and moderate to severe disease should be considered with caution due to limited evidence and in some instances wide confidence intervals suggesting wider variability in the estimated point estimate.

Quality of life

There is insufficient evidence to assess whether one topical calcineurin inhibitor has "better" quality of life profile compared with another topical calcineurin inhibitor. Indirect assessment between the topical agents was also difficult due to varied methods of reporting quality of life information. In general, patients randomized to tacrolimus or pimecrolimus reported improvements in quality of life scores relative to patients randomized to vehicle.

Active-control trials with topical steroids in moderate to severe disease

Tacrolimus compared with lowest-potency topical steroid (class 7)

Tacrolimus 0.03% and 0.1% ointments were significantly more effective by week 3 in achieving treatment success than hydrocortisone acetate 1% ointment in children 2 to 15 years of age.

Tacrolimus compared with mid- to low-potency topical steroids (class 5-6)
In adults, at 3 weeks hydrocortisone-17-butyrate 0.1% ointment was more effective than tacrolimus 0.03% ointment but was only as effective as tacrolimus 0.1% ointment. Use of hydrocortisone butyrate was not restricted from the head-and-neck region. In a 24-week study, tacrolimus 0.1% ointment was more effective than hydrocortisone acetate 1% ointment (for head/neck or intertriginous application) plus hydrocortisone butyrate 0.1% ointment (for trunk/limbs) in adults.

Tacrolimus ointment compared with medium-potency topical steroids (class 4)
In children (2 to 15 years) with underlying moderate to severe disease who have acute severe to very severe disease symptoms, treatment with methylprednisolone aceponate 0.1% ointment applied once daily was as effective as tacrolimus 0.03% ointment applied 2 times a day over 3 weeks.

Pimecrolimus 1% cream compared with mid-potency topical steroids (class 5)
Betamethasone-17-valerate 0.1% cream (applied to all areas except head/neck) was more effective than pimecrolimus 1% cream in adults at the end of 3 weeks. Triamcinolone acetonide 0.1% cream plus hydrocortisone acetate 1% cream (applied to head/neck and/or intertriginous areas) was as effective as pimecrolimus 1% cream in adults over 52 weeks.

Maintenance or prevention (24 to 52 weeks)

No head-to-head studies assessed long-term outcomes on maintenance or preventative therapy with tacrolimus or pimecrolimus. None of the tacrolimus trials were long in duration and none evaluated long-term outcomes; therefore, indirect comparative assessments could not be conducted. Only 5 pimecrolimus vehicle-controlled trials were studied up to 52 weeks and assessed outcomes such as rebound flare-ups. Of these, 4 vehicle-controlled trials showed that pimecrolimus 1% cream was significantly more effective than vehicle in preventing flares and reducing topical steroid use in patients with mild to severe disease over 24 to 52 weeks. Two of these studies reported “time to first flare” and found that pimecrolimus was more effective than vehicle (53 to 144 days to first flare compared with 13 to 26 days). One trial reported no significant difference between pimecrolimus and vehicle for the percentage of days on which patients’ required topical steroid use.

Detailed Assessment

Shorter-term treatment (≤12 weeks)

Twelve studies[16, 17, 22-26, 34-38] were short in duration (2 head-to-head studies, 5 tacrolimus vehicle-controlled trials, 5 pimecrolimus vehicle-controlled trials). Of these, 3 publications[16, 22, 35] were pooled analyses of unpublished trials that were also found in the FDA Medical and Statistical Reviews.

Table 5. Head-to-head studies 6 weeks in duration

Author, year Quality	Population	N (total)	Disease severity	Intervention	Comparison
Kempers 2004[34] Fair	Children, 2 to 17 years	139	Moderate 99%	Tacrolimus 0.03% ointment twice daily	Pimecrolimus 1% cream twice daily
Paller 2005-(a)[35] Fair	Children, 2 to 15 years	423	Mild 100%	Tacrolimus 0.03% ointment twice daily	Pimecrolimus 1% cream twice daily
Paller 2005-(b)[35] Fair	Children, 2 to 15 years	224	Moderate 75%-81% Severe 16%-21%	Tacrolimus 0.1% ointment twice daily	Pimecrolimus 1% cream twice daily
Paller 2005-(c)[35] Fair	Adults, >16 years	413	Mild 31%-33% Moderate 44%-47% Severe 16%-19%	Tacrolimus 0.1% ointment twice daily	Pimecrolimus 1% cream twice daily

Table 6. Tacrolimus: Vehicle-controlled trials 3 to 12 weeks in duration

Author, year Quality	Population	N (tacrolimus)	N (total)	Disease severity	Intervention
Boguniewicz 1998[23] Fair	Children, 7 to 16 years	0.03%: 43 0.1%: 49	180	Moderate 73%-88% Severe 12%-27%	Tacrolimus 0.03%, 0.1%, 0.3% ointment twice daily
Hanifin 2001[16] (study #35) Fair	Adults, >16 years	0.03%: 103 0.1%: 99	304	Moderate 39%-52% Severe 48%-61%	Tacrolimus 0.03%, 0.1% ointment twice daily
Hanifin 2001[16] (study #36) Fair	Adults, >16 years	0.03%: 108 0.1%: 110	328	Moderate 36%-43% Severe 56%-64%	Tacrolimus 0.03%, 0.1% ointment twice daily
Paller 2001[17] Fair	Children, 2 to 15 years	0.03%: 117 0.1%: 118	351	Moderate 36%-41% Severe 60%-64%	Tacrolimus 0.03%, 0.1%, ointment twice daily
Ruzicka 1997[24] Fair	Children and adults, 13 to 60 years	0.03%: 54 0.1%: 54	213	Moderate to severe (% not reported)	Tacrolimus 0.03%, 0.1%, 0.3% ointment twice daily
Schachner 2005[37] Fair	Children, 2 to 15 years	0.03%: 158	317	Mild 60% Moderate 40%	Tacrolimus 0.03% ointment twice daily

Table 7. Pimecrolimus: Vehicle-controlled trials 3 to 12 weeks in duration

Author, year Quality	Population	n (pimecrolimus)	N (total)	Disease severity	Intervention
Eichenfield 2002[22] (study #305) Fair	Children, 1-17 years	130	198	Mild 22%-27% Moderate 56%-64%	Pimecrolimus 1% cream twice daily
Eichenfield 2002[22] (study #307) Fair	Children, 1-17 years	137	205	Mild 37%-38% Moderate 57%-59%	Pimecrolimus 1% cream twice daily
Ho 2003[36] Fair	Infants, 3-23 months	123	186	Mild 33% Moderate 67%	Pimecrolimus 1% cream twice daily
Kaufmann 2004[25] Fair	Infants, 3-23 months	129	195	Moderate 58%-59% Severe 26%	Pimecrolimus 1% cream twice daily
Luger 2001[a, 26] Fair	Adults, ≥18 years	45	260	Moderate 91%-95% Severe 5%-9%	Pimecrolimus 1% cream twice daily
Van Leent 1998[38] Fair	Adults	Twice daily: 16	34	Moderate (ADSI ~7/15)	Pimecrolimus 1% cream once or twice daily

Abbreviations: ADSI, atopic dermatitis severity index.
[a] This trial also included an active-control, betamethasone-17-valerate 0.1% cream.

Mild to moderate disease

We did not identify any head-to-head studies or vehicle-controlled trials of tacrolimus 0.1% ointment or pimecrolimus 1% cream (applied twice daily) in patients with mild to moderate disease.

Direct evidence

Two head-to-head studies compared tacrolimus 0.03% ointment with pimecrolimus 1% cream in children with mild to moderate disease (Table 5, Evidence Tables 3 and 4).[34, 35] At the end of 6 weeks, there was no significant difference in the proportion of patients achieving treatment success between those using tacrolimus 0.03% ointment or pimecrolimus 1% cream (pooled relative risk, 1.19, 95% CI 0.98 to 1.45). For both treatment groups, atopic dermatitis cleared or nearly cleared in almost 50% of patients (pooled rate, 46%, 95% CI 40.0% to 52.0%).

Pruritus scores could not be pooled for meta-analysis, but both studies showed no significant difference in itching between patients randomized to tacrolimus 0.03% ointment or pimecrolimus 1% cream. One study[34] showed 70% of tacrolimus-treated patients reporting a pruritus severity score ≤1, which indicates absent or mildly severe itchiness, compared with 64% of pimecrolimus-treated patients, P=0.49. The other study[35] reported the change in pruritus score from baseline using a 10-cm visual analog scale. Numerically and statistically tacrolimus was shown to be more effective at reducing pruritus however, the clinical relevance of an estimated between-group difference of -0.5 cm change is unclear (change in score from baseline: tacrolimus, -2.9 cm, P≤0.01 and pimecrolimus, -2.4 cm, P≤0.01).

None of the head-to-head studies reported patient assessment of their disease control and none stratified efficacy outcomes by body location (for example, intertriginous areas or head/neck region) or by percent of affected body surface area. Only 1 pooled analysis of 3 unpublished trials reported efficacy results for patients with head/neck involvement.[35] Of 1060

patients across 3 trials, 710 patients with mild to severe disease had head/neck involvement. In this subgroup, a larger proportion of tacrolimus-treated patients experienced improvements in their EASI scores from baseline than pimecrolimus-treated patients at the end of 6 weeks (tacrolimus, 57% improvement compared with pimecrolimus 42%, *P*=0.01).

Indirect evidence

The evidence base for indirect comparison between tacrolimus 0.03% ointment and pimecrolimus 1% cream in mild to moderate disease is limited to 4 vehicle-controlled trials: 1 tacrolimus and 3 pimecrolimus (Tables 6 and 7, Evidence Tables 5 to 8).[22, 36, 37] Trials were 6 weeks in duration and included infants (3 to 23 months) and children (2 to 15 years). By the end of the study, there was no significant difference between tacrolimus or pimecrolimus in achieving treatment success per IGA score ≤1 (pooled relative risk, 0.97, 95% CI 0.63 to 1.48). A sensitivity analysis excluding 1 pimecrolimus trial[36] which may have contributed to high heterogeneity (I^2= 86.2%) was conducted. This pimecrolimus trial found larger treatment differences in the percent achieving treatment success in younger infants (3 months to 1 year of age) than in infants and children >1 year in age (65.5% compared with 46.3%). Omitting the 1 pimecrolimus study did not change the initial conclusion (pooled relative risk 1.05, 95% CI 0.64 to 1.72) (Table 8, Appendix E).

One pooled analysis[39] of 2 tacrolimus trials was not included in our indirect meta-analysis because of duplicate results with Schachner 2005. The combined results in the pooled publication indicated that tacrolimus was superior to vehicle (Evidence Tables 7 and 8).

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)

Trial	Mean patient age	Tacrolimus 0.03% (n= 158)	Pimecrolimus 1% (n= 390)	Vehicle (n= 358)
Schachner 2005[37]	6.9 years	50.6%	N/A	25.8%
Ho 2003[36]	13 months	N/A	54.5%	23.8%
Eichenfield 2002[22] study #305	6.7 years	N/A	37.7%	16.2%
Eichenfield 2002[22] study #307	6.8 years	N/A	32.1%	20.6%
Pooled rate		50.6%	41.3%	22.1%
95% CI		(42.8%–58.4%)	(28.4%–54.2%)	(17.8%–26.5%)
Heterogeneity statistics	Cochrane's Q	N/A	14.5 (*P*<0.001)	3.1 (*P*=0.38)
	I^2	N/A	86.2%	2.8%
Pooled relative risk (95% CI): 0.97 (0.63–1.48)				

Because pruritus outcomes were reported differently across 4 vehicle-controlled trials, indirect meta-analysis was not performed. In all trials however, treatment with tacrolimus 0.03%

ointment or pimecrolimus 1% cream was more effective than vehicle in reducing pruritus (Evidence Tables 5 to 8 for individual trial results).

Because none of the tacrolimus 0.03% trials reported patient evaluation of overall disease control, no comparative evidence for this outcome is available. In all 3 pimecrolimus vehicle-controlled trials,[22, 36] patients randomized to use pimecrolimus reported "good/complete" or "better/much better" resolution of their disease than patients using vehicle (range: pimecrolimus 60% to 71.5% compared with vehicle 27% to 40%, $P<0.05$).

None of the 4 vehicle-controlled trials stratified outcomes by percent of affected body surface area, and only 1 tacrolimus vehicle-controlled trial[37] and 1 pimecrolimus pooled analysis[22] reported mean EASI score improvement in the head/neck region in a subgroup of patients. For tacrolimus, a 55.2% improvement in mean EASI score was observed compared with a 0% improvement with vehicle. For pimecrolimus, a 58% improvement in mean EASI score was observed compared with a 3.9% worsening in disease control with vehicle.[22, 37] It appears that pimecrolimus 1% cream may be slightly more or as effective as tacrolimus 0.03% ointment in treating atopic dermatitis of the head and neck. More evidence is needed to verify the findings.

Active-control trials with topical steroids in mild to moderate disease

We did not find any active-control trials that investigated tacrolimus 0.03% or 0.1% ointment or pimecrolimus 1% cream with topical steroids in populations with mild to moderate atopic dermatitis.

Moderate to severe disease

Direct evidence

We did not find head-to-head studies comparing lower strength tacrolimus (0.03%) ointment with pimecrolimus 1% cream (both applied twice daily) in infants, children, or adults with moderate to severe disease.

Higher strength tacrolimus 0.1% ointment was found to be more effective than pimecrolimus 1% cream (both applied twice daily) in 1 small head-to-head study.[35] In 224 children, 32.4% of tacrolimus-treated patients achieved treatment success compared with 17.7% of pimecrolimus-treated patients (relative risk 1.83, 95%, CI 1.13 to 2.96). A post hoc subgroup analysis of 281 adults also showed similar findings (tacrolimus, 40.4% compared with pimecrolimus, 22.1%, $P=0.001$).[40] The adults had moderate to severe disease and were part of an original trial of 431 adults.

Pruritus scores from 1 head-to-head study and 1 post hoc subgroup analysis suggest that tacrolimus 0.1% ointment may be as effective or slightly better than pimecrolimus 1% cream in relieving itch at the end of 6 weeks.[35, 40] Patients randomized to tacrolimus exhibited a -3.7 cm change in pruritus score compared with a -2.0 cm change with pimecrolimus, $P\leq0.01$.[35] Although statistical significance was reached favoring tacrolimus, clinical significance between the groups is unclear (estimated between-group difference: -1.7 cm). Moreover, a subgroup analysis showed little difference in the improvement of pruritus score between those using tacrolimus or pimecrolimus (tacrolimus, -3.5 cm compared with pimecrolimus, -2.9 cm).[40]

None of the head-to-head studies evaluated patient assessment of overall disease control or stratified efficacy outcomes by percent of affected body surface area. One post hoc subgroup analysis of 193 adults with head/neck involvement showed a 66% improvement with tacrolimus compared with a 49% improvement with pimecrolimus in reducing signs and symptoms of atopic dermatitis on the face and neck ($P=0.02$ for between-group difference).[40]

Indirect evidence

Seven trials comparing tacrolimus 0.03% or 0.1% ointment or pimecrolimus 1% cream with vehicle over 3 to 12 weeks in patients with moderate to severe disease were identified.[16, 17, 23-26] Each trial showed that treatment with tacrolimus or pimecrolimus was superior to vehicle in achieving treatment success.

For our indirect meta-analysis we were limited to 1 small pimecrolimus trial and 3 small tacrolimus trials (2 of the tacrolimus trials were reported in a pooled analysis[16]). Results from 3 other trials[23, 24, 26] could not be included in the indirect analysis due to heterogenous outcome reporting. It should be noted that, among the trials included in our analysis, the proportion of patients with severe disease in the pimecrolimus vehicle-controlled trials was slightly different from the proportion of patients with severe disease in the tacrolimus vehicle-controlled trials (Tables 6 and 7).

Tacrolimus 0.03%

No statistically significant difference in treatment success was found when tacrolimus 0.03% ointment was indirectly compared with pimecrolimus 1% cream in 4 vehicle-controlled trials (pooled relative risk, 0.89, 95% CI 0.38 to 2.07; Tables 6, 7, and 9).[16, 17, 25] The change in EASI score was also not significantly different between the treatment groups (pooled weighted mean difference 3.77, 95% CI -1.93 to 9.47); however, this result should be interpreted with caution due to wide confidence interval.

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)

Trial	Duration (weeks)	Mean age (years)	Tacrolimus 0.1% (n= 327)	Tacrolimus 0.03% (n= 328)	Pimecrolimus 1% (n= 129)	Vehicle (n= 394)
Paller 2001[17]	12	6.1	40.7%	35.9%	N/A	6.9%
Kaufmann 2004[25]	4	12.0	N/A	N/A	53.5%	10.6%
Hanifin 2001[16] study #35	12	38.6	35.4%	29.1%	N/A	7.8%
Hanifin 2001[16] study #36	12	38.5	38.2%	25.9%	N/A	5.5%
Pooled rates			38.2%	30.2%	53.5%	7.0%
95% CI			(32.9%–43.4%)	(24.4%–36.0%)	(44.9%–62.1%)	(4.5%–9.6%)
Heterogeneity statistics		Cochrane's Q	0.7 (*P*=0.72)	2.7 (*P*=0.25)	N/A	1.5 (*P*=0.68)
		I^2	0%	26.9%	N/A	0%
Tacrolimus 0.03% compared with pimecrolimus: pooled relative risk (95% CI): 0.89 (0.38 to 2.07)						
Tacrolimus 0.1% compared with pimecrolimus: pooled relative risk (95% CI): 1.17 (0.43 to 3.20)						

There was also little difference between tacrolimus 0.03% ointment and pimecrolimus 1% cream in the improvement of pruritus (pooled weighted mean difference, 0.86, 95% CI -0.69 to 2.41)[16, 17, 25] or in proportion of patients reporting "good/complete" or "better/much better" assessment of their disease control over 3 to 12 weeks (pooled relative risk, 0.98, 95% CI 0.56 to 1.73).[16, 25, 26]

Only 3 small trials reported treatment results specific to the head/neck region.[23-25] Again, because efficacy outcomes were reported differently, we could not pool the findings. The 3 trials showed that use of tacrolimus or pimecrolimus was more effective than vehicle in reducing EASI scores from baseline (Evidence Tables 5 to 8).

Tacrolimus 0.1%

No significant difference was found in rates for achieving treatment success when higher strength tacrolimus 0.1% was indirectly compared with pimecrolimus 1% cream in infants and children over 4 to 12 weeks (pooled relative risk, 1.17, 95% CI 0.43 to 3.20; Tables 6, 7, and 9).[17, 25] When results from 2 additional tacrolimus trials (conducted in adults) were combined with the meta-analysis above, the overall conclusion remained unchanged while the confidence interval narrowed (pooled relative risk 1.12, 95% CI 0.48 to 2.57).[16]

When pruritus scores for tacrolimus 0.1% ointment were indirectly compared with pimecrolimus 1%, minimal difference between the groups was found (pooled weighted mean difference for the change in score, 0.74, 95% CI -0.83 to 2.31).[16, 17, 25] Four trials reported patient assessment of "overall disease control," and the pooled results also showed no significant difference between tacrolimus and pimecrolimus (pooled relative risk 1.07, 95% CI 0.53 to 2.13).[16, 25, 26]

Two tacrolimus and 1 pimecrolimus vehicle-controlled trial reported efficacy results for the head/neck region.[23-25] Due to heterogeneous outcome reporting we could not pool the results. In all 3 publications, tacrolimus 0.1% ointment or pimecrolimus 1% cream had superior efficacy in the treatment of atopic dermatitis in the head/neck areas relative to vehicle (Evidence Tables 5 to 8).

Quality of life

The evidence base for quality of life assessment is limited to 1 pooled tacrolimus study,[41] 5 pimecrolimus studies,[42-46] and 1 pooled pimecrolimus study.[47] Only 2 pimecrolimus studies reported quality of life information up to 12 months[46, 47] while the remaining studies reported quality of life assessments at 3 and 12 weeks. None of the studies were combined for quantitive analysis due to varied reporting of outcomes (for example, some studies reported change in quality score while others reported percent improvement in score).

In 1 pooled study of 2 tacrolimus vehicle-controlled trials,[41] patients treated with tacrolimus observed significant quality of life benefit compared with patients treated with vehicle at the end of 12 weeks ($P<0.05$). In this study, 3 different tools for measuring quality of life were used in adults, children, and toddlers: the Dermatology Life Quality Index (DLQI) questionnaire for adults, the Children Dermatology Life Quality Index (CDLQI) survey for children, and a modified CDLQI tool for caregivers of toddlers. More than 50% of patients had moderate to severe disease at baseline.

For pimecrolimus, 1 short-term and long-term study showed significant improvements in quality of life with the use of pimecrolimus 1% cream than with vehicle in patients with mild to moderate disease.[44, 47] In a 20-week, open-label extension study, however, no significant

difference in quality of life scores was found between patients treated with pimecrolimus during a 6 week double blind phase compared with patients who switched from vehicle to pimecrolimus during the open-label period (change in PIQoL-AD score, 76.1% compared with 77.1%).[46] Four different quality of life tools were used in pimecrolimus trials: the Parent's Index of Quality of Life Atopic Dermatitis (PIQoL-AD), the Quality of Life Index Atopic Dermatitis (QoL-AD), the CDLQI, and the DLQI.

Active-control trials with topical steroids in moderate to severe disease

Tacrolimus 0.03% compared with lowest-potency topical steroid (class 7)
In 2 trials, treatment with tacrolimus 0.03% ointment was significantly more effective than treatment with hydrocortisone acetate 1% ointment applied twice daily for 3 weeks in children ages 2 to 15 years (Evidence Tables 9 and 10).[19, 48] One study reported percent improvement in modified EASI score (using mean area under the curve) from baseline (55.2% compared with 36.0%, $P<0.001$) and percent of patients with a PGE of 90% to 100% improvement (38.5% compared with 15.7%, $P=0.001$).[19] The other study reported median percent improvement in modified EASI score (78.7% compared with 47.2%, $P<0.001$) and the percent of patients reporting "much better/better" overall disease (82.9% compared with 50.7%, *P* value not reported).[48] Only 1 trial provided data on patient assessment (10-cm scale) of pruritus;[48] it showed an approximately -3.5 cm change in pruritus score for tacrolimus 0.03% compared with a -2.0 cm change for hydrocortisone (estimated between-group difference, -1.5 cm).

Tacrolimus 0.1% compared with lowest-potency topical steroid (class 7)
Children (2 to 15 years) treated with tacrolimus 0.1% ointment for 3 weeks showed greater improvement in their modified EASI scores than patients treated with hydrocortisone acetate 1% ointment (Evidence Tables 9 and 10).[19] The percent improvement in modified EASI using mean AUC was 60.2% for tacrolimus compared with 36.0% for hydrocortisone acetate, $P<0.001$. The proportion of patients who achieved "clear or almost clear" resolution of disease was also larger for tacrolimus-treated patients than those on hydrocortisone acetate (48.4% compared with 15.7%, $P=0.001$).

Tacrolimus 0.03% compared with mid- to low-potency steroid (class 5-6)
In adults hydrocortisone-17-butyrate 0.1% ointment (applied to head/neck and trunk/limbs twice daily) was more effective than tacrolimus 0.03% ointment applied twice daily for 3 weeks (Evidence Tables 9 and 10).[18] The percent improvement in modified EASI based on mean AUC for hydrocortisone butyrate 0.1% and tacrolimus 0.03% was 63.9% and 53.0%, respectively, $P=0.002$. The proportion of patients with a PGE of 90% to 100% improvement was 51.4% hydrocortisone butyrate and 37.6% tacrolimus, $P<0.05$. Changes in pruritus score were not reported, although this outcome was measured.

Tacrolimus 0.1% compared with mid- to low-potency steroids (class 5-6)
At 3 weeks, hydrocortisone-17-butyrate 0.1% ointment was as effective as tacrolimus 0.1% ointment in adults (Evidence Tables 9 and 10).[18] Percent improvement in modified EASI based on mean-AUC for hydrocortisone butyrate 0.1% and tacrolimus 0.03% was 63.9% compared with 63.5%, $P>0.5$. The proportion of patients with PGE of 90% to 100% improvement was also similar between the 2 treatment groups: 51.4% for hydrocortisone butyrate and 49.2% for tacrolimus, $P>0.5$. Topical steroid was used on head/neck and trunk/limbs regions.

One study showed tacrolimus 0.1% ointment was more effective than a combination of hydrocortisone butyrate 0.1% ointment (for trunk and limbs) and hydrocortisone acetate 1% ointment (for head and neck) applied twice daily in adults for 24 weeks.[49] More than 70% of tacrolimus-treated patients improved by at least 60% per modified EASI score compared with approximately 50% of topical steroid-treated patients, $P<0.001$. Similarly, atopic dermatitis cleared in more patients using tacrolimus (61.3%) than using topical steroid combination (46.4%, $P<0.001$). Patient assessment of pruritus was not reported, but the authors state that pruritus improved "substantially for patients in both treatment groups." Patients more often than not reported their overall disease control as "much better/better" with tacrolimus (86.6%) than with steroids (71.8%, $P<0.001$). A significant proportion of patients receiving topical steroids withdrew from the study due to lack of efficacy (42.1% compared with 25.5%); it is unclear whether the withdrawal rate due to lack of efficacy was influenced by potential "unequal" dose comparisons between hydrocortisone acetate 1% ointment and tacrolimus 1% ointment for use in head/neck region in those with moderate to severe disease.

Pimecrolimus 1% compared with mid-potency topical steroids (class 5)

One fair-quality placebo- and active-control trial evaluated pimecrolimus 1% cream and betamethasone-17-valerate 0.1% cream in adults with moderate to severe disease (Evidence Tables 11 and 12).[26] Betamethasone-17-valerate 0.1% cream, applied to all affected areas except head/neck, showed significantly greater improvement than pimecrolimus on adjusted EASI score (approximately 80% compared with 45%) and in relief of pruritus to a severity score ≤1 (81.0% compared with 46.7%) compared with pimecrolimus 1% cream at the end of week 3. Patients treated with betamethasone-17-valerate also reported higher rates of "moderately clear or better" (>50%) improvement in disease from baseline than patients treated with pimecrolimus 1% cream (88.1% compared with 53.3%, *P* value not reported).

In a 52-week study, treatment with pimecrolimus 1% cream was as effective as treatment with twice daily applications of triamcinolone acetonide 0.1% cream (applied to trunk/limbs) and hydrocortisone acetate 1% cream (applied to head/neck and intertriginous areas) in adults with moderate to severe disease.[50] Clinical improvement (EASI and IGA) and time to first remission were similar between the 2 groups (time to first remission pimecrolimus 221 days compared with topical steroids 212 days). Time to first recurrence of symptoms was also similar (pimecrolimus 14 days compared with topical steroids 17 days). These results should be considered with caution since "completer" population was analyzed instead of intention to treat population. More than 35% of pimecrolimus-treated patients withdrew due to lack of efficacy which could have affected the magnitude of treatment effect between the groups.

Tacrolimus 0.03% compared with medium-potency steroid (class 4)

In children (2 to 15 years) with underlying moderate to severe disease who had an IGA score ≥4 during the trial, methylprednisolone aceponate 0.1% ointment applied *once a day* for 3 weeks was as effective as tacrolimus 0.03% ointment applied *twice daily* (Evidence Tables 9 and 10).[51] The proportion for which PGE showed 90% to 100% improvement differed by 0.3% between the 2 treatment groups (95% CI -11.1% to -11.5%). By the end of the study, dermatitis had resolved or nearly resolved in about 67% of children. Mean percentage change from baseline in EASI score did not differ between treatment groups (tacrolimus 85.3% compared with methylprednisolone 89.7%, $P=0.067$). The only statistically significant difference between the treatment groups was for patient assessment of pruritus (100 mm visual analog scale). Change

from baseline for tacrolimus was -49.8 mm compared with methylprednisolone -61.7 mm, *P*=0.0004.

Maintenance or prevention (24 to 52 weeks)

The evidence base for long-term maintenance or prevention of atopic dermatitis in infants, children, or adults with mild to severe disease is lacking. None of the tacrolimus trials were longer than 12 weeks and none assessed outcomes such as time to first flare, percent of patients without flares, or percent of patients requiring topical steroid rescue. Only 5 long-term trials comparing pimecrolimus 1% cream with vehicle at 24 and 52 weeks were identified (Evidence Tables 3 and 4).[42, 45, 52-54] Of these, 1 trial was conducted in adults[52] while the remaining trials were in infants and children. Three trials included patients with mild to moderate disease[42, 45, 53] and 2 trials included patients with moderate to severe disease.[52, 54] The primary objectives in these studies was to determine whether early treatment with pimecrolimus 1% cream would prevent progression to acute flares and whether pimecrolimus 1% cream exhibited topical steroid-sparing effect (that is, decreased topical steroid use).

In 4 of 5 trials, all patients were required to use emollients and were instructed to apply pimecrolimus 1% cream or vehicle twice daily at the first signs or symptoms of atopic dermatitis (erythema, pruritus, etc).[42, 45, 52, 53] Treatment with study medications was to continue until signs and symptoms of disease cleared. "Rescue" topical steroids (moderate in potency) were initiated in situations where study medication was inadequate (for example, acute flare). Treatment with topical steroids continued up to the maximum duration allowed based on package labeling or determined by each country involved in the trials. It is important to note that after acute flares were treated with a "rescue" topical steroid, *residual disease* was treated for another 7 consecutive days with pimecrolimus 1% cream or vehicle.

Criteria for determining an acute flare were specified in all 5 trials; however, there was some variation in the definitions. Two trials[42, 45] defined acute flare as IGA score ≥4 plus the use of topical steroid within 3 days of physician assessment. One study[52] defined flare as the need for topical steroid use for more than 3 consecutive days. One study defined flare based on symptoms,[54] and one study defined flare as IGA score ≥4.[53]

Only 1 trial reported whether patients were screened for topical steroid response prior to study enrollment.[54] This study was conducted in children with predominately moderate to severe disease, and prior to enrollment patients were treated with a moderately potent topical steroid (prednicarbate 0.25% cream twice daily) for at least 7 days (maximum 21 days). If during the run-in phase skin conditions did not improve with prednicarbate cream, patients were excluded from participating. This study did not require patients to use emollients but allowed emollient use on an "as needed" or "if needed" basis.

Topical steroids that were used in these studies were difluprednate 0.02% cream, prednicarbate 0.25% cream, hydrocortisone butyrate 0.1% cream, clobetasone butyrate 0.05% cream, triamcinolone acetonide 0.02% cream, fluticasone propionate 0.05% cream, mometasone furoate 0.1% cream, and hydrocortisone valerate 0.2% cream.

Results from 4 of 5 trials support that pimecrolimus 1% cream was more effective than vehicle (applied twice daily) in preventing flares and minimizing steroid use in patients with mild to severe disease. The proportion of pimecrolimus-treated patients and vehicle-treated patients without flares at the end of 24 and 52 weeks was 51% to 68% compared with 28% to 34% across the trials,[42, 45, 52, 53] and the percent requiring topical steroid rescue was 35% to 51% compared with 63% to 78%.[42, 45, 52] Two of these trials showed that pimecrolimus was more

effective at delaying the time to first flare than vehicle (53 to 144 days compared with 13 to 26 days), and 1 study reported that pimecrolimus-treated patients used less topical steroid than vehicle-treated patients (percent number of days on topical steroids 14.2% compared with 37.2%).[52, 53] One trial,[54] however, reported conflicting results for the percentage of days on which patients required topical steroid use. No difference was found between pimecrolimus 1% cream and vehicle at the end of 24 weeks (29%±25% of days compared with 35%±25% of days, *P*=0.18). The authors attribute statistical insignificance to inconsistent grading of disease severity at baseline: Some investigators in this trial classified patients as having severe disease per Rajka and Langeland criteria, when according to the static IGA method, severity would have been considered more mild or moderate. Consequently, more patients with mild to moderate disease were enrolled in the study than patients with severe disease. The authors further explained that when patients with mild to moderate disease were excluded in a post hoc analysis, statistical difference was found between pimecrolimus and vehicle. This suggests that patients with severe disease may not have responded as well to prednicarbate as patients with mild and moderate disease, and thus the addition of pimecrolimus helped achieve statistical significance in this group. The results from this trial should be considered with caution and should also be verified in larger prospective long-term trials.

Key Question 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 depending on location of application (for example, head and neck, flexures, hands, feet, intertriginous regions), depending on body surface area involved, or depending on treatment duration?

Summary

Good-quality long-term studies evaluating serious harms between tacrolimus and pimecrolimus are still lacking. One fair-quality, short-term, nested case-control study suggests that the odds of lymphoma associated with tacrolimus and pimecrolimus are low for patients who had up to 4 years exposure to these agents.

Total withdrawal rate (pooled relative risk 0.81, 95% CI 0.63 to 1.05) and rate of withdrawal due to adverse events (pooled relative risk 0.50, 95% 0.16 to 1.54) were not significantly different between tacrolimus (0.03%, 0.1%) ointment and pimecrolimus 1% cream in patients with varying disease severity included in 2 head-to-head studies. However, it appears that tacrolimus-treated patients were less likely to withdraw from therapy than pimecrolimus-treated patients at the end of 6 weeks (pooled relative risk 0.39, 95% CI 0.21 to 0.72). Indirect meta-analysis showed similar findings for total withdrawal and withdrawal due to adverse events. Total withdrawal rates were slightly higher or similar for patients using tacrolimus (0.03%, 0.1%) ointment or pimecrolimus 1% cream compared with patients using topical steroids in 4 active-control trials.

Application site reactions were the most common skin-related events reported. However, the incidence of burning, stinging, erythema, and irritation did not differ significantly between tacrolimus (0.03%, 0.1%) ointment and pimecrolimus 1% cream. When compared with vehicle,

a higher proportion of tacrolimus- and pimecrolimus-treated patients reported application site reactions such as burning and stinging (49% to 52% compared with 29% to 35%). When compared with topical steroids, significantly more tacrolimus- and pimecrolimus-treated patients reported application site reactions.

There is insufficient direct evidence regarding serious viral skin infections between tacrolimus and pimecrolimus. Most trials reported cases or rates of viral skin infections but not all studies were consistent in the reporting, suggesting potential for selective reporting bias. Therefore, interpretation of results for serious viral skin infections should be considered with some caution.

None of the studies reported skin atrophy, telangiectasia, or adrenal suppression. Only 1 study reported 3 cases of skin striae (0.9%) in adults receiving triamcinolone acetonide 0.1% cream (trunk/limbs) with hydrocortisone acetate 1% cream (head/neck) over 52 weeks. The extent and severity of striae however, were not described.

Detailed Assessment

Harms

Good-quality long-term studies evaluating serious harms-related events were not found. Most trials generally reported total withdrawal rates and adverse events that emerged during the study duration. Methods for collecting data on harms (such as actively querying patients rather than passively allowing patients to report events) were not explicitly stated, and methods on how adverse events were reported (for example, selective reporting) generally were not detailed. In a few publications, the number of cases of particular viral infections was not reported for each treatment group, making comparisons between topical calcineurin inhibitors difficult. The severity of adverse events was also not predefined or prespecified. Only 3 trials prespecified investigation of application site reactions, local skin infections, and cases of acne.[34, 37, 50] General adverse events were collected from 25 vehicle- and active-control trials.[16-19, 23, 24, 34, 35, 39, 48, 49, 51, 54, 55 22, 25, 26, 36, 42, 45, 50, 52, 53, 56, 57] Case-control studies were also included, but due to limited good-quality evidence on serious harms, the results from these studies should be interpreted with caution.

Lymphomas

No good-quality, long-term, comparative studies assessing serious harms were found. Two short-term nested case-control studies were identified[58, 59] (Evidence Tables 13 and 14).

In one study, cases of lymphoma and controls were identified from the PharMetrics database which included 73 United States health plans. Records in this database represented the managed care population. Initially, patients with ICD-9 codes for atopic dermatitis were collected from 1995 to 2005 and 502 283 patients were identified. After excluding patients who had <6 months enrollment in the database, who had an existing diagnosis of lymphoma, cancer, HIV, or AIDS, or a history of immunosuppressive therapy or transplantation, a total of 293 253 patients served as the final cohort for analysis. Of these patients, 75% were enrolled in the database from 2001 (limiting the total duration of exposure to topical calcineurin inhibitors). At the index date (day an ICD-9 code was present for atopic dermatitis), 1.5% to 3% of patients were using topical pimecrolimus or topical tacrolimus. In these patients, 14 and 11 cases of lymphoma were identified for those exposed to topical pimecrolimus and topical tacrolimus.

ICD-9 codes were used to identify cases of lymphoma and the cases were reviewed by blinded hematologists. The authors of this nested case-control state that misclassification of events could have easily occurred and thus the results should be considered with caution.[58] Based on these findings, the odds of lymphoma associated with topical pimecrolimus (odds ratio 0.82, 95% CI 0.42 to 1.61) and with topical tacrolimus (odds ratio 0.79, 95% CI 0.37 to 1.71) indicated a low overall rate of lymphoma.

A second study that assessed the risk of nonmelanoma skin cancer (NMSC) in adults with "dermatitis" who used topical calcineurin inhibitors compared with those who did not use these topical agents was rated poor (see Evidence tables 13 and 14).[59] This study was rated poor-quality based on a combination of factors which included: high risk of recall bias which is also the most difficult to control; unclear description of the selection of the sample; no information regarding duration of illness or duration of exposure to topical calcineurin inhibitors; and no explanation on how missing histologic data (used to confirm cases of NMSC) were handled.

Skin atrophy, telangiectasia, adrenal suppression, or skin striae

None of the included studies reported skin atrophy, telangiectasia, or adrenal suppression. Of the 2 long-term active-control trials, only 1 study[50] reported 3 cases of skin striae (0.9%) in adult patients receiving triamcinolone acetonide 0.1% cream (trunk/limbs) with hydrocortisone acetate 1% cream (head/neck) over 52 weeks. Striae were identified on legs between 4 and 9 months and remained until the end of the study. The extent and severity of striae were not described. In this trial, the mean percentage of days on which patients needed to apply study medication was 83.4% in the topical steroid group and 88.7% in the pimecrolimus group. The median percentage of days of exposure to study medications was 95.6% in the topical steroid group compared with 99.5% in the pimecrolimus group. The study did not report baseline disease duration, did not report baseline topical steroid usage (including potency of past topical steroids and duration of use), and did not specify whether all patients were screened for evidence of skin changes prior to study enrollment.

Withdrawals

Total withdrawal (pooled relative risk 0.81, 95% CI 0.63 to 1.05) and withdrawal due to adverse events (pooled relative risk 0.50, 95% CI 0.16 to 1.54) did not significantly differ for tacrolimus (0.03%, 0.1%) ointment or pimecrolimus 1% cream in patients with mild to severe disease included in 2 short-term head-to-head studies.[34, 35] However, tacrolimus-treated patients were less likely to withdraw from treatment secondary to lack of efficacy than pimecrolimus-treated patients (pooled rate 2.2% compared with pooled rate 6.5%; pooled relative risk 0.39, 95% CI 0.21 to 0.72). Indirect meta-analysis of pimecrolimus and tacrolimus vehicle-controlled trials also showed no significant differences for total withdrawal rates (pooled relative risk, 0.81, 95% CI 0.61 to 1.07) and for withdrawals due to adverse events (pooled relative risk, 0.63, 95% CI 0.27 to 1.46) across populations in 14 trials.[16, 17, 22-26, 36, 39, 42, 45, 52-54]

When tacrolimus 0.03%, 0.1%, and pimecrolimus 1% was compared with vehicle alone, significantly more patients randomized to vehicle withdrew from the trials (pooled rates: vehicle 41% compared with topical calcineurin inhibitors 18%).[16, 17, 22-26, 36, 39, 42, 45, 52-54] The most common reason for withdrawal for vehicle-treated patients was due to lack of efficacy (pooled rates: 28% compared with 6.8% for topical calcineurin inhibitors).[16, 17, 22, 23, 25, 26, 36, 39, 42, 45, 52, 53]

Of the 4 active-control trials where topical steroids were as effective as or more effective than tacrolimus[18, 51] or pimecrolimus,[26, 50] the rates of total withdrawal were numerically less for

those randomized to topical steroids (Table 10). One study,[50] however, did not report withdrawal rates for both treatment groups. For the remaining active-control trials[19, 48, 49] where tacrolimus was shown to be more effective than topical steroids, total withdrawal rates were greater for patients on topical steroids (pooled rates: tacrolimus 14% compared with topical steroids 23%). The most common reason was lack of efficacy (pooled rates: tacrolimus 4.6% compared with topical steroids 11%).

Table 10. Total withdrawal rates for 4 active-control trials

	Tacro	Pime	BMV	HB	MPA	TC+HCA
Luger 2001[26]	---	15.5%	7.1%	---	---	---
Luger 2004[50]	---	41.2%	---	---	---	Not reported
Reitamo 2002[18]	11.5%	---	---	9.1%	---	---
Bieber 2007[51]	4.4%	---	---	---	1.6%	---

Abbreviations: BMV, betamethasone valerate; HB, hydrocortisone butyrate; HCA, hydrocortisone acetate; MPA, methylprednisolone aceponate; pime, pimecrolimus; tacro, tacrolimus; TC, triamcinolone acetonide.

Application site reactions

Commonly reported adverse events were application site reactions (burning, stinging, pruritus, etc). Head-to-head studies found no significant difference between tacrolimus (0.03%, 0.1%) ointment and pimecrolimus 1% cream for rates of burning and stinging (pooled relative risk 0.96, 95% CI 0.44 to 2.08) or for erythema and irritation (pooled relative risk 0.58, 95% CI 0.18 to 1.85).[34, 35]

Across vehicle-controlled trials, significantly more tacrolimus-treated (up to 52%) and pimecrolimus-treated patients (up to 49%) experienced burning, stinging, erythema, or irritation during treatment compared with up to 35% of vehicle-treated patients.[16, 17, 22-26, 36, 39, 42, 45, 52-55]

In 7 active-control trials, patients randomized to tacrolimus (0.03%, 0.1%) ointment or pimecrolimus 1% cream experienced greater incidence of application site reactions compared with topical steroids (range across trials: tacrolimus 20% to 52% compared with topical steroids 7% to 15%; pimecrolimus 46% to 49% compared with 10% to 24% topical steroids).[18, 19, 26, 48-51]

Herpes simplex virus, molluscum contagiosum, eczema herpeticum, herpes zoster

Not all trials consistently reported these adverse events and it is unknown whether the reporting of these adverse events was done without bias; therefore, interpretation of the magnitude of frequency should be considered with some caution. Table 11 reports the number of cases of serious or potentially serious skin infection events gathered from 10 trials[17, 25, 39, 42, 45, 52-55, 57] and 7 active-control trials over 3 to 52 weeks.[18, 19, 35, 48-51] There were 2 cases of herpes simplex dermatitis observed with pimecrolimus, of which 1 case was considered to be study medication related. Antiviral treatment was required for this patient.[45] A case of Kaposi's varicelliform eruption together with a bacterial skin infection (possibly related) in 1 patient on tacrolimus 0.03% ointment, twice daily was also identified.[48]

Table 11. Number of cases collected from 17 trials (N=7761) that reported serious viral infections

	Vehicle-controlled (n=3461)			Active-control (n=4300)		
	Tacro	Pime	Vehicle	Tacro	Pime	Topical steroids
Herpes simplex virus	24	15	13	20	18	40
Molluscum contagiosum	8	3	1	NR	0	2
Eczema herpeticum	5	3	4	0	2	0
Herpes zoster	7	NR	0	NR	1	2

Abbreviations: NR, not reported; pime, pimecrolimus; tacro, tacrolimus.

A retrospective cohort study in 388 Japanese patients suggested that there may be little difference between topical tacrolimus and topical steroids in the incidence of herpes simplex infection or eczema herpeticum of the face or neck. However, the duration of topical steroid and topical tacrolimus exposure was significantly different between the groups. Patients were exposed to 1 to 3 years of tacrolimus 0.1% ointment to face and neck after 1 to 17 years' exposure to low-and medium potency topical steroids. The rate of herpes simplex infection was 6.1 cases/100 patient-years during topical steroid exposure compared with 6.9 cases/100 patient-years during topical tacrolimus exposure (P=0.52) while the rate of eczema herpeticum was 3.1 cases/100 patient-years compared with 2.9 cases/100 patient-years between the groups (P=0.96).[60]

In an open-label, single-arm study of 799 adults and children, the reported rates of herpes simplex (6.3%), molluscum contagiosum (2.0%), eczema herpeticum (0.8%), and herpes zoster (3.3%) were considered low in patients treated with topical tacrolimus over approximately 3 years.[61] Interpretation of the results should be considered with some caution since about 50% of patient data on adverse events were not available for inclusion. Of the 50% who withdrew from the open-label study, 37% withdrew because of noncompliance, patient withdrawal, and loss to follow-up.

Key Question 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?

Summary

There is insufficient comparative evidence in subgroup populations based on age, gender, race, and comorbidities for tacrolimus and pimecrolimus. Most subgroup analyses were performed for either tacrolimus or pimecrolimus in vehicle-controlled trials. All subgroup analyses were conducted post hoc.

Detailed Assessment

None of the head-to-head studies conducted subgroup analyses. Subgroup analyses conducted in all the placebo trials were done post hoc. We did not find any good-or fair-quality studies evaluating the use of topical calcineurin inhibitors in patients with eyelid dermatitis.

Age

In 1 pimecrolimus trial, infants 3 months to 1 year of age exhibited larger treatment effect in the proportion of patients with IGA score ≤1 than those who were 1 to 2 years of age (65.5% compared with 46.3%).[36]

Ethnic origin

One post hoc analysis suggests that black adults (N=110) had better response in achieving >90% improvement of disease with tacrolimus 0.1% ointment (29.1% compared with vehicle, 7%, *P*=0.002) than with tacrolimus 0.03% ointment (16.4% compared with vehicle 7%, *P*=0.112) applied twice daily.[16]

One post hoc analysis (N=589) showed no difference between white and multiracial patients in their response to pimecrolimus 1% cream (vehicle-corrected value in percent patients with IGA score ≤1: white 21.4% and multiracial 20.6%, *P*>0.5).[62] The multiracial group included: 41.8% black, 11.6% Asian, and 46.6% Hispanic patients.

Baseline disease severity

Patients with mild disease who were randomized to tacrolimus 0.03% ointment exhibited pronounced improvement in treatment success (tacrolimus 56.7% compared with vehicle 32.3%, *P*=0.0007) than in patients with moderate disease (tacrolimus 41% compared with vehicle 15.9%, *P*=0.001).[37]

In 241 patients with severe disease, treatment with tacrolimus 0.1% ointment was more effective than with tacrolimus 0.03% ointment (both applied twice daily).[16] The success rate (>90% improvement based on PGE) was 35.0% for tacrolimus 0.1% ointment and was 19.5% for tacrolimus 0.03% ointment (*P*=0.009).

One study showed little difference in the treatment effect of pimecrolimus 1% cream in patients with mild and moderate disease compared with patients with severe disease (estimated between-group difference in % change in EASI score: 4.7%).[25]

Body surface area involved with atopic dermatitis

In a small subset of adults with >75% body surface area affected by atopic dermatitis (N=82), patients receiving tacrolimus 0.1% ointment were more likely to achieve treatment success than patients receiving tacrolimus 0.03% ointment applied twice daily (30.2% compared with 5.1%, *P*=0.004) at the end of 12 weeks.[16]

Chronic hand dermatitis

Overall, the use of pimecrolimus 1% cream 2 times a day plus vinyl glove dressing in adults with chronic hand dermatitis was more effective than vehicle in achieving an IGA score ≤1 at 3 weeks.[56]

Assumptions and Limitations

Pooling across populations and stratifying by disease severity

Based on epidemiologic studies and the natural history of the disease, we assumed that adult patients included in our review did not have drastically different disease (were not treatment resistant) when compared with infants and children, such that data could not be pooled. In all trials, information regarding treatment resistance or duration of illness in adults was not specified at baseline. Therefore, we assumed no significant differences in the populations and stratified the results by disease severity prior to analyses. Subsequent to making this assumption, results from a subgroup in 1 pimecrolimus study[36] suggested that there may be some differences in treatment effect between infants younger than 1 years of age and infants and children > 1 years of age or adults. Further studies that include a broader range of patients (by age) need to be conducted to verify these findings in both tacrolimus and pimecrolimus trials.

We assumed that baseline disease severity as reported in the trials accurately represented clinical appearance, regardless of the type of tool used to evaluate severity. This is a limitation of our review, since we know that not all assessment tools on severity have been validated and not all tools measure similar items, but currently there is no standard method for assessing disease severity.[20] For example, in 1 trial,[54] investigators using the Rajka and Langeland method graded disease as severe, whereas according to the IGA system the disease was considered more mild. However, most of the trials included in our meta-analyses used the IGA system to grade baseline disease severity.

Assessment of outcomes

To determine the success of treatment we combined the proportion of patients with IGA score ≤1 with the proportion of patients with a PGE improvement score of 90% to 100%. This method is a limitation, because IGA scoring is based on descriptive appearance of the disease made at each visit, whereas PGE is scored by comparing the extent of improvement of the skin relative to baseline. Based on these differences in the assessment of atopic dermatitis, the magnitude of treatment success using IGA score may be slightly larger than with results using PGE.

Another limitation to our review is that IGA and PGE scores are not fully validated assessment tools.[20] Currently, 3 assessment tools (SCORAD, EASI, NESS) have been validated in multiple studies. None of the trials in this review reported efficacy outcomes using SCORAD or NESS methods. EASI scores were reported in most studies but outcome reporting was heterogeneous across trials (some trials reporting percent improvement in score and others reporting with change in score). Where possible, we included EASI results in our review to support the findings from IGA and PGE.

Generalizability characteristics

Many of the included studies did not provide sufficient baseline information regarding disease duration, prior treatment failures, and comorbidities thus limiting the generalizability of results to broader populations. More than two-thirds of the included trials were conducted in children and infants with 10% to 40% of their total body surface area affected by atopic dermatitis with more than 70% with head/neck involvement. Fifty to 70% of patients were white and female while 20% to 30% of patients were black. Patients with concomitant infections or significant comorbid conditions (for example, Netherton syndrome) were excluded from the trials. Patients were

recruited mostly from dermatology or allergy clinics and were likely managed by specialists (for example, dermatologists).

SUMMARY

Table 12 summarizes the definitions used for terms used in grading the strength of evidence and Appendix F describes how we assessed the overall strength of evidence. Table 13 summarizes results of this review.

Table 12. Definitions of overall strength of evidence

Grade	Definition
High	**High confidence that the evidence reflects the true effect.** Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	**Moderate confidence that the evidence reflects the true effect.** Further research may change our confidence in the estimate of effect and may change the estimate.
Low	**Low confidence that the evidence reflects the true effect.** Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate.
Insufficient	Evidence either is unavailable or does not permit estimation of an effect.

Table 13. Summary of the evidence by key question

	Strength of evidence	Conclusion
Key Question 1. Efficacy/effectiveness		
Shorter-term treatment (≤12 weeks)		
Mild to moderate disease	Direct evidence: Moderate-high—for comparing tacrolimus 0.03% ointment with pimecrolimus cream 1% Insufficient—for comparing tacrolimus 0.1% ointment with pimecrolimus 1% cream Indirect evidence: Moderate—for comparing tacrolimus 0.03% ointment with pimecrolimus 1% cream	Tacrolimus 0.03% ointment was as effective as pimecrolimus 1% cream in treating atopic dermatitis in patients with mild to moderate disease in 2 head-to-head studies (pooled relative risk, 1.19, 95% CI 0.98-1.45) and an indirect meta-analysis of 4 vehicle-controlled trials (pooled relative risk 0.97, 95% CI 0.63-1.48). Improvements in pruritus were also not significantly different between tacrolimus 0.03% ointment and pimecrolimus 1% cream in 2 head-to-head studies. None of the studies in mild to moderate disease reported patients' assessment of overall disease control and none stratified efficacy outcomes depending on affected body surface area. Evidence evaluating treatment effect in the head/neck area was found but is limited. Only 1 tacrolimus trial and 1 pooled pimecrolimus study reported mean EASI score improvement which suggests that pimecrolimus 1% cream may be slightly more, or as effective as tacrolimus 0.03% ointment in the head/neck region.

	Strength of evidence	Conclusion
		We did not find any studies that investigated higher strength tacrolimus 0.1% ointment with pimecrolimus 1% cream in patients with mild to moderate disease. No active-control studies comparing tacrolimus (0.03%, 0.1%) ointment with topical steroids in this population were identified.
Moderate to severe disease	Direct evidence: Insufficient—for comparing tacrolimus 0.03% ointment with pimecrolimus 1% cream Moderate-low—for comparing tacrolimus 0.1% ointment with pimecrolimus 1% cream Indirect evidence: Low—for tacrolimus 0.03% and 0.1% ointment with pimecrolimus 1% cream	There is insufficient head-to-head evidence comparing lower strength tacrolimus 0.03% ointment with pimecrolimus 1% cream in patients with moderate to severe disease. Indirect comparison of tacrolimus 0.03% ointment and pimecrolimus 1% cream showed no statistically significant difference in achieving treatment success at the end of 3 to 12 weeks (pooled relative risk 0.89, 95% CI 0.38 to 2.07). There was also little difference in pruritus score (pooled weighted mean difference 0.86, 95% CI -0.69 to 2.41) or patient assessment of overall disease control (pooled relative risk 0.98, 95% CI 0.56 to 1.73) between the topical agents. Direct and indirect evidence reported conflicting results when higher strength tacrolimus 0.1% ointment was compared with pimecrolimus 1% cream in patients with moderate to severe disease. Tacrolimus 0.1% ointment was shown to be more effective in achieving treatment success than pimecrolimus 1% cream (relative risk 1.83, 95% CI 1.13 to 2.96) and maybe slightly better at relieving pruritus (-3.7 cm compared with -2.0 cm, $P \leq 0.01$) at 6 weeks in a small head-to-head study. In contrast, indirect comparison of 3 tacrolimus and 1 pimecrolimus trial revealed no statistically significant differences in treatment success (pooled relative risk 1.17, 95% CI 0.43 to 3.20), change in pruritus score (pooled weighted mean difference 0.74, 95% CI -0.83 to 2.31), or differences in patients' assessment of overall disease control (pooled relative risk 1.07, 95% CI 0.53 to 2.13). There is insufficient evidence to assess whether one topical calcineurin inhibitor has "better" quality of life profile compared with another topical calcineurin inhibitor. Indirect assessment between the topical agents was also difficult due to varied methods of reporting quality of life information. In general, patients randomized to tacrolimus or pimecrolimus reported improvements in quality of life scores relative to patients randomized to vehicle.

	Strength of evidence	Conclusion
Key Question 1. Efficacy/effectiveness		
Maintenance or prevention (24 to 52 weeks)		
	Direct evidence: Insufficient	No head-to-head studies assessed long-term outcomes on maintenance or preventative therapy with tacrolimus or pimecrolimus.
	Indirect evidence: Insufficient—(only pimecrolimus trials available)	None of the tacrolimus trials were long in duration and none evaluated long-term outcomes; therefore, indirect comparative assessments could not be conducted. Only 5 pimecrolimus vehicle-controlled trials were studied up to 52 weeks in duration and assessed outcomes such as rebound flare-ups. Of these, 4 vehicle-controlled trials showed that pimecrolimus 1% cream was significantly more effective than vehicle in preventing flares and reducing topical steroid use in patients with mild to severe disease over 24 to 52 weeks. Two of these studies reported "time to first flare" and found that pimecrolimus was more effective than vehicle (53 to 144 days to first flare compared with 13 to 26 days). One trial reported no significant difference between pimecrolimus and vehicle for the percentage of days on which patients' required topical steroid use.
Key Question 2. Harms	**Strength of evidence**	**Conclusion**
	Direct evidence: Moderate—for commonly reported adverse events Insufficient for rare or serious adverse events	Good-quality long-term studies evaluating serious harms between tacrolimus and pimecrolimus are still lacking. One fair-quality, short-term, nested case-control study suggests that the odds of lymphoma associated with tacrolimus and pimecrolimus are low for patients who had up to 4 years exposure to these agents.
	Indirect evidence: Moderate—for commonly reported adverse events Low—for rare or serious adverse events	Total withdrawal rate (pooled relative risk 0.81, 95% CI 0.63 to 1.05) and rate of withdrawal due to adverse events (pooled relative risk 0.50, 95% 0.16 to 1.54) were not significantly different between tacrolimus (0.03%, 0.1%) ointment and pimecrolimus 1% cream in 2 head-to-head studies. However, it appears that tacrolimus-treated patients were less likely to withdraw from therapy than pimecrolimus-treated patients at the end of 6 weeks (pooled relative risk 0.39, 95% CI 0.21 to 0.72). Indirect meta-analysis showed similar findings for total withdrawal and withdrawal due to adverse events. Total withdrawal rates were slightly higher or similar for patients using tacrolimus or pimecrolimus compared with patients using topical steroids in 4 active-control trials.

	Strength of evidence	Conclusion
		Application site reactions were the most common skin-related events reported. However, the incidence of burning, stinging, erythema, and irritation did not differ significantly between tacrolimus (0.03%, 0.1%) ointment and pimecrolimus 1% cream. When compared with vehicle, a higher proportion of tacrolimus- and pimecrolimus-treated patients reported application site reactions such as burning and stinging (49% to 52% compared with 29% to 35%). When compared with topical steroids, significantly more tacrolimus- and pimecrolimus-treated patients reported application site reactions. There is insufficient or limited direct and indirect evidence regarding serious viral skin infections between tacrolimus and pimecrolimus. Most trials reported cases or rates of viral skin infections but not all studies were consistent in the reporting, suggesting potential for selective reporting bias. Therefore, interpretation of results for serious viral skin infections should be considered with some caution. None of the studies reported skin atrophy, telangiectasia, or adrenal suppression. Only 1 study reported 3 cases of skin striae (0.9%) in adults receiving triamcinolone acetonide 0.1% cream (trunk/limbs) with hydrocortisone acetate 1% cream (head/neck) over 52 weeks. The extent and severity of striae however, were not described.
Key Question 3. Subgroups	**Strength of evidence**	**Conclusion**
	Low	There is insufficient comparative evidence in subgroup populations based on age, gender, race, and comorbidities for tacrolimus and pimecrolimus. Most subgroup analyses were performed for either tacrolimus or pimecrolimus in vehicle-controlled trials. All subgroup analyses were conducted post hoc. For pimecrolimus 1% cream, more infants (3 months to 1 year) achieved IGA score ≤1 than infants and children >1 year. 65.5% compared with 46.3% (1 trial). Black adults had better treatment response (>90% improvement) with tacrolimus 0.1% than tacrolimus 0.03% ointment (1 post hoc analysis) No significant differences in treatment effect between white and multiracial (black, Asian, Hispanic) patients using pimecrolimus 1% cream was found (1 post hoc analysis). Larger treatment effects were seen with tacrolimus

Strength of evidence	Conclusion
	0.03% ointment in patients with mild disease than with moderate disease (1 trial). Tacrolimus 0.1% appeared to be more effective in patients with severe disease than tacrolimus 0.03% ointment (1 study). In patients with >75% of their body surface affected with atopic dermatitis, tacrolimus 0.1% ointment appeared to be more effective than tacrolimus 0.03% ointment (1 study). Pimecrolimus 1% cream plus occlusive dressing was more effective than vehicle in treating chronic hand dermatitis (1 trial).

REFERENCES

1. Williams H. How epidemiology has challenged three prevailing concepts about atopic dermatitis. *Journal of Allergy and Clinical Immunology.* 2006;118(1):219-213.

2. Illi S, Mutius EV, Lau S, Nickel R, Gruber C, Niggemann B. The natural course of atopic dermatitis from birth to age 7 years and the association with asthma. 2004.

3. Leung D, Y.M., Bieber T. Atopic Dermatitis. *Lancet.* 2003;361:151-160.

4. Bieber T. Mechanisms of disease: Atopic Dermatitis. *New England Journal of Medicine.* 2008;358(14):1483-1494.

5. Fitzpatrick. *Fitzpatrick's Dermatology in General Medicine-Chap 14*. 7 ed: McGraw-Hill.

6. Leung DYM, Nicklas RA, Li JT, Bernstein IL, Blessing-Moore J. Disease management of atopic dermatitis: an updated practice parameter. *Annals of Allergy, Asthma and Immunology.* 2004;93:S1-S21.

7. Oranje AP, Glazenburg EJ, Wolkerstorfert A, de Waard-van der Spek FB. Practical issues on interpretation of scoring atopic dermatitis: the SCORAD index, objective SCORAD and the three-item severity score. *British Journal of Dermatology.* 2007;157:645-648.

8. Van der Meer JB, Glazenburg EJ, Mulder PGH, Coenraads PJ. The management of moderate to severe atopic dermatitis in adults with topical fluticasone propionate. *British Journal of Dermatology.* 1999;140:1114-1121.

9. Berth-Jones J, Damstra RJ, Golsch S, et al. Twice weekly fluticasone propionate added to emollient maintenance treatment to reduce risk of relapse in atopic dermatitis: randomised, double blind, parallel group study. *British Medical Journal.* 2003;326:1367-1374.

10. Anonymous. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. *CRD Report.* 2001(4).

11. Harris RP, Helfand M, Woolf SH. Current methods of the US Preventive Services Task Force: a review of the process. . *American Journal of Preventive Medicine.* 2001;20((3 Suppl)):21-35.

12. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods of Meta-Analysis in Medical Research*: John Wiley & Sons, Inc.; 2000.

13. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technology Assessment (Winchester, England).* 2005 Jul 2005;9(26):1-134.

14. Higgins JP, Thompson SG, Deeks JJ, Altman DJ. Measuring consistency in meta analysis. *British Medical Journal.* 2003;327(7414):557-560.

15. Higgins JPT, Thompson S, G. Quantifying heterogeneity in meta analysis. *Stat Med.* 2002;21(11):1539-1558.

16. Hanifin JM, Ling MR, Langley R, Breneman D, Rafal E. Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: part I, efficacy. *J Am Acad Dermatol.* Jan 2001;44(1 Suppl):S28-38.

17. Paller A, Eichenfield LF, Leung DY, Stewart D, Appell M. A 12-week study of tacrolimus ointment for the treatment of atopic dermatitis in pediatric patients. *J Am Acad Dermatol.* Jan 2001;44(1 Suppl):S47-57.

18. Reitamo S, Rustin M, Ruzicka T, et al. Efficacy and safety of tacrolimus ointment compared with that of hydrocortisone butyrate ointment in adult patients with atopic dermatitis. *Journal of Allergy and Clinical Immunology.* 2002;109(3):547-555.
19. Reitamo S, Van Leent EJM, Ho V, et al. Efficacy and safety of tacrolimus ointment compared with that of hydrocortisone acetate ointment in children with atopic dermatitis. *Journal of Allergy and Clinical Immunology.* 2002;109(3):539-546.
20. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema: A systematic review. *Journal of Allergy & Clinical Immunology.* 2007;120(6):1389-1398.
21. Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol.* Feb 2001;10(1):11-18.
22. Eichenfield LF, Lucky AW, Boguniewicz M, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. *Journal of the American Academy of Dermatology.* Apr 2002;46(4):495-504.
23. Boguniewicz M, Fiedler VC, Raimer S, Lawrence ID, Leung DY, Hanifin JM. A randomized, vehicle-controlled trial of tacrolimus ointment for treatment of atopic dermatitis in children. Pediatric Tacrolimus Study Group. *Journal of Allergy & Clinical Immunology.* Oct 1998;102(4 Pt 1):637-644.
24. Ruzicka T, Bieber T, Schopf E, et al. A short-term trial of tacrolimus ointment for atopic dermatitis. European Tacrolimus Multicenter Atopic Dermatitis Study Group. *New England Journal of Medicine.* Sep 18 1997;337(12):816-821.
25. Kaufmann R, Folster-Holst R, Hoger P, et al. Onset of action of pimecrolimus cream 1% in the treatment of atopic eczema in infants. *Journal of Allergy & Clinical Immunology.* Nov 2004;114(5):1183-1188.
26. Luger T, Van Leent EJ, Graeber M, et al. SDZ ASM 981: an emerging safe and effective treatment for atopic dermatitis. *British Journal of Dermatology.* Apr 2001;144(4):788-794.
27. Hung S-H, Lin Y-T, Chu C-Y, et al. Staphylococcus colonization in atopic dermatitis treated with fluticasone or tacrolimus with or without antibiotics. *Annals of Allergy, Asthma, & Immunology.* Jan 2007;98(1):51-56.
28. Schnopp C, Remling R, Mohrenschlager M, Weigl L, Ring J, Abeck D. Topical tacrolimus (FK506) and mometasone furoate in treatment of dyshidrotic palmar eczema: a randomized, observer-blinded trial. *Journal of the American Academy of Dermatology.* Jan 2002;46(1):73-77.
29. Torok HM, Maas-Irslinger R, Slayton RM. Clocortolone pivalate cream 0.1% used concomitantly with tacrolimus ointment 0.1% in atopic dermatitis. *Cutis.* Aug 2003;72(2):161-166.
30. Ashcroft D, Chen L-C, Garside R, Stein K, Williams H. Topical pimecrolimus for eczema. *Cochrane Database of Systematic Reviews.* 2007;4:4.
31. Garside R, Stein K, Castelnuovo E, et al. The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England).* 2005 Jul 2005;9(29):iii.

32. Iskedjian M, Piwko C, Shear NH, Langley RGB, Einarson TR. Topical calcineurin inhibitors in the treatment of atopic dermatitis: a meta-analysis of current evidence. *American Journal of Clinical Dermatology.* 2004;5(4):267-279.
33. Ashcroft DM, Dimmock P, Garside R, Stein K, Williams HC. Efficacy and tolerability of topical pimecrolimus and tacrolimus in the treatment of atopic dermatitis: meta-analysis of randomised controlled trials. *BMJ.* Mar 5 2005;330(7490):516.
34. Kempers S, Boguniewicz M, Carter E, et al. A randomized investigator-blinded study comparing pimecrolimus cream 1% with tacrolimus ointment 0.03% in the treatment of pediatric patients with moderate atopic dermatitis. *Journal of the American Academy of Dermatology.* Oct 2004;51(4):515-525.
35. Paller AS, Lebwohl M, Fleischer AB, Jr., et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. *Journal of the American Academy of Dermatology.* May 2005;52(5):810-822.
36. Ho VC, Gupta A, Kaufmann R, et al. Safety and efficacy of nonsteroid pimecrolimus cream 1% in the treatment of atopic dermatitis in infants. *Journal of Pediatrics.* Feb 2003;142(2):155-162.
37. Schachner LA, Lamerson C, Sheehan MP, et al. Tacrolimus Ointment 0.03% Is Safe and Effective for the Treatment of Mild to Moderate Atopic Dermatitis in Pediatric Patients: Results From a Randomized, Double-Blind, Vehicle-Controlled Study. *Pediatrics.* September 1, 2005 2005;116(3):e334-342.
38. Van Leent EJ, Graber M, Thurston M, Wagenaar A, Spuls PI, Bos JD. Effectiveness of the ascomycin macrolactam SDZ ASM 981 in the topical treatment of atopic dermatitis. *Archives of Dermatology.* Jul 1998;134(7):805-809.
39. Chapman MS, Schachner LA, Breneman D, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. *Journal of the American Academy of Dermatology.* 2005;53(2, Supplement 2):S177-S185.
40. Fleischer AB, Jr., Abramovits W, Breneman D, Jaracz E, group USCtos. Tacrolimus ointment is more effective than pimecrolimus cream in adult patients with moderate to very severe atopic dermatitis. *Journal of Dermatological Treatment.* 2007;18(3):151-157.
41. Drake L, Prendergast M, Maher R, et al. The impact of tacrolimus ointment on health-related quality of life of adult and pediatric patients with atopic dermatitis. *J Am Acad Dermatol.* Jan 2001;44(1 Suppl):S65-72.
42. Kapp A, Papp K, Bingham A, et al. Long-term management of atopic dermatitis in infants with topical pimecrolimus, a nonsteroid anti-inflammatory drug. *Journal of Allergy & Clinical Immunology.* Aug 2002;110(2):277-284.
43. Leo HL, Bender BG, Leung SB, Tran ZV, Leung DYM. Effect of pimecrolimus cream 1% on skin condition and sleep disturbance in children with atopic dermatitis. *Journal of Allergy & Clinical Immunology.* Sep 2004;114(3):691-693.
44. Staab D, Kaufmann R, Brautigam M, Wahn U, Group CAC-S. Treatment of infants with atopic eczema with pimecrolimus cream 1% improves parents' quality of life: a multicenter, randomized trial. *Pediatric Allergy & Immunology.* Sep 2005;16(6):527-533.
45. Wahn U, Bos JD, Goodfield M, et al. Efficacy and safety of pimecrolimus cream in the long-term management of atopic dermatitis in children. *Pediatrics.* Jul 2002;110(1 Pt 1):e2.

46. Whalley D, Huels J, McKenna SP, Van Assche D. The benefit of pimecrolimus (Elidel, SDZ ASM 981) on parents' quality of life in the treatment of pediatric atopic dermatitis. *Pediatrics.* Dec 2002;110(6):1133-1136.
47. McKenna SP, Whalley D, de Prost Y, et al. Treatment of paediatric atopic dermatitis with pimecrolimus (Elidel, SDZ ASM 981): impact on quality of life and health-related quality of life. *Journal of the European Academy of Dermatology & Venereology.* Mar 2006;20(3):248-254.
48. Reitamo S, Harper J, Bos JD, et al. 0.03% tacrolimus ointment applied once or twice daily is more efficacious than 1% hydrocortisone acetate in children with moderate to severe atopic dermatitis: results of a randomized double-blind controlled trial. *British Journal of Dermatology.* 2004;150(3):554-562.
49. Reitamo S, Ortonne JP, Sand C, et al. A multicentre, randomized, double-blind, controlled study of long-term treatment with 0·1% tacrolimus ointment in adults with moderate to severe atopic dermatitis. *British Journal of Dermatology.* 2005;152(6):1282-1289.
50. Luger TA, Lahfa M, Folster-Holst R, et al. Long-term safety and tolerability of pimecrolimus cream 1% and topical corticosteroids in adults with moderate to severe atopic dermatitis. *Journal of Dermatological Treatment.* Jun 2004;15(3):169-178.
51. Bieber T, Vick K, Folster-Holst R, et al. Efficacy and safety of methylprednisolone aceponate ointment 0.1% compared to tacrolimus 0.03% in children and adolescents with an acute flare of severe atopic dermatitis. *Allergy.* Feb 2007;62(2):184-189.
52. Meurer M, Folster-Holst R, Wozel G, et al. Pimecrolimus cream in the long-term management of atopic dermatitis in adults: a six-month study. *Dermatology.* 2002;205(3):271-277.
53. Siegfried E, Korman N, Molina C, Kianifard F, Abrams K. Safety and efficacy of early intervention with pimecrolimus cream 1% combined with corticosteroids for major flares in infants and children with atopic dermatitis.[erratum appears in J Dermatolog Treat. 2006;17(4):256]. *Journal of Dermatological Treatment.* 2006;17(3):143-150.
54. Zuberbier T, Heinzerling L, Bieber T, Schauer U, Klebs S, Brautigam M. Steroid-sparing effect of pimecrolimus cream 1% in children with severe atopic dermatitis. *Dermatology.* 2007;215(4):325-330.
55. Soter NA, Fleischer AB, Webster GF, Monroe E, Lawrence I. Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: Part II, Safety. *Journal of the American Academy of Dermatology.* 2001;44(1, Part 2):S39-S46.
56. Belsito DV, Fowler JF, Jr., Marks JG, Jr., et al. Pimecrolimus cream 1%: a potential new treatment for chronic hand dermatitis. *Cutis.* Jan 2004;73(1):31-38.
57. Ling M, Gottlieb A, Pariser D, et al. A randomized study of the safety, absorption and efficacy of pimecrolimus cream 1% applied twice or four times daily in patients with atopic dermatitis. *Journal of Dermatological Treatment.* Aug 2005;16(3):142-148.
58. Arellano FM, Wentworth CE, Arana A, Fernandez C, Paul CF. Risk of lymphoma following exposure to calcineurin inhibitors and topical steroids in patients with atopic dermatitis. *Journal of Investigative Dermatology.* Apr 2007;127(4):808-816.
59. Margolis DJ, Hoffstad O, Bilker W. Lack of association between exposure to topical calcineurin inhibitors and skin cancer in adults. *Dermatology.* 2007;214(4):289-295.

60. Hashizume H, Yagi H, Ohshima A, et al. Comparable risk of herpes simplex virus infection between topical treatments with tacrolimus and corticosteroids in adults with atopic dermatitis. *British Journal of Dermatology.* Jun 2006;154(6):1204-1206.
61. Hanifin JM, Paller AS, Eichenfield L, et al. Efficacy and safety of tacrolimus ointment treatment for up to 4 years in patients with atopic dermatitis. *J Am Acad Dermatol.* Aug 2005;53(2 Suppl 2):S186-194.
62. Eichenfield LF, Lucky AW, Langley RGB, et al. Use of pimecrolimus cream 1% (Elidel) in the treatment of atopic dermatitis in infants and children: the effects of ethnic origin and baseline disease severity on treatment outcome. *International Journal of Dermatology.* Jan 2005;44(1):70-75.

Appendix A. Search strategy

Database: Ovid MEDLINE(R) <1950 to November Week 2 2007>
Search Strategy:
---
1 (topical$ adj5 (tacrolimus or pimecrolimus or calcineurin inhibitor$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word] (619)
2 protopic.mp. (62)
3 ascomycin macrolactam.mp. (19)
4 pimecrolimus.mp. (414)
5 elidel.mp. (58)
6 ASM 981.mp. (47)
7 (topical$ adj2 FK506).mp. (35)
8 1 or 2 or 3 or 4 or 5 or 6 or 7 (877)
9 limit 8 to (humans and english language) (741)
10 limit 9 to yr="1990 - 2008" (741)
11 from 10 keep 1-741 (741)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <4th Quarter 2007>
Search Strategy:
---
1 protopic.mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (13)
2 elidel.mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (14)
3 ascomycin macrolactam.mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (6)
4 (topical$ adj5 (tacrolimus or pimecrolimus or calcineurin inhibitor$)).mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (61)
5 ASM 981.mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (45)
6 (topical$ adj2 FK506).mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (3)
7 1 or 2 or 3 or 4 or 5 or 6 (118)
8 from 7 keep 1-118 (118)

Database: EBM Reviews - Cochrane Database of Systematic Reviews <4th Quarter 2007>
Search Strategy:
---
1 protopic.mp. [mp=title, abstract, full text, keywords, caption text] (1)
2 elidel.mp. [mp=title, abstract, full text, keywords, caption text] (1)
3 ascomycin macrolactam.mp. [mp=title, abstract, full text, keywords, caption text] (0)
4 (topical$ adj5 (tacrolimus or pimecrolimus or calcineurin inhibitor$)).mp. [mp=title, abstract, full text, keywords, caption text] (12)
5 ASM 981.mp. [mp=title, abstract, full text, keywords, caption text] (1)

6 (topical$ adj2 FK506).mp. [mp=title, abstract, full text, keywords, caption text] (0)
7 1 or 2 or 3 or 4 or 5 or 6 (12)
8 from 7 keep 1-12 (12)

Database: EBM Reviews - Database of Abstracts of Reviews of Effects <4th Quarter 2007>
Search Strategy:

---

1 protopic.mp. [mp=title, full text, keywords] (0)
2 elidel.mp. [mp=title, full text, keywords] (0)
3 ascomycin macrolactam.mp. [mp=title, full text, keywords] (0)
4 (topical$ adj5 (tacrolimus or pimecrolimus or calcineurin inhibitor$)).mp. [mp=title, full text, keywords] (4)
5 ASM 981.mp. [mp=title, full text, keywords] (0)
6 (topical$ adj2 FK506).mp. [mp=title, full text, keywords] (0)
7 1 or 2 or 3 or 4 or 5 or 6 (4)
8 from 4 keep 1-4 (4)

Appendix B. Quality assessment of drug class reviews for the Drug Effectiveness Review Project

Study quality is objectively assessed using predetermined criteria for internal validity, based on a combination of the US Preventive Services Task Force and the National Health Service Centre for Reviews and Dissemination[1, 2] criteria.

All included studies are assessed for quality and assigned a rating of "good", "fair", or "poor". Studies that have a fatal flaw are rated poor-quality. A fatal flow is reflected by failure to meet combinations of criteria that may be related in indicating the presence of bias. An example would be failure or inadequate procedures for randomization and/or allocation concealment combined with important differences in prognostic factors at baseline. Studies that meet all criteria are rated good-quality and the remainder is 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 true difference between the compared drugs.

Systematic Reviews[3, 4]	
1. Were the search methods reported? *Were the search methods used to find evidence (original research) on the primary questions stated?* **"Yes" if the review states the databases used, date of most recent searches, and some mention of search terms.**	The purpose of this index is to evaluate the scientific quality (adherence to scientific principles) of research overviews (review articles) published in the medical literature. It is not intended to measure literary quality, importance, relevance, originality, or other attributes of overviews. The index is for assessing overviews of primary ("original") research on pragmatic questions regarding causation, diagnosis, prognosis, therapy, or prevention. A research overview is a survey of research. The same principles that apply to epidemiological surveys apply to overviews: a question must be clearly specified, a target population identified and accessed, appropriate information obtained from that population in an unbiased fashion, and conclusions derived, sometimes with the help of formal statistical analysis, as is done in "meta-analyses". The fundamental difference between overviews and epidemiological studies is the unit of analysis, not the scientific issues that the questions in this index address. Since most published overviews do not include a methods section, it is difficult to answer some of the questions in the index. Base your answers, as much as possible, on information provided in the overview. If the methods that were
2. Was the search comprehensive? *Was the search for evidence reasonably comprehensive?* **"Yes" if the review searches at least 2 databases and looks at other sources (such as reference lists, hand searches, queries experts).** *Note: EMBASE was launched in 1972 and CDSR was launched in 1994; therefore, papers prior to 1994 can be graded "Yes" if only one database is searched.*	
3. Were the inclusion criteria reported? *Were the criteria used for deciding which studies to include in the overview reported?*	

Systematic Reviews[3, 4]	
4. Was selection bias avoided? *Was bias in the selection of studies avoided?* **"Yes" if the review reports how many studies were identified by searches, numbers excluded, and gives appropriate reasons for excluding them (usually because of pre-defined inclusion/exclusion criteria).**	used are reported incompletely relative to a specific question, score it as "can't tell", unless there is information in the overview to suggest either the criterion was or was not met.
5. Were the validity criteria reported? *Were the criteria used for assessing the validity of the included studies reported?*	
6. Was validity assessed appropriately? *Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?* **"Yes" if the review reports validity assessment and did some type of analysis with it (e.g. sensitivity analysis of results according to quality ratings, excluded low-quality studies, etc.)**	
7. Were the methods used to combine studies reported? *Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?* **"Yes" for studies that did qualitative analysis if there is some mention that quantitative analysis was not possible and reasons that it could not be done, or if 'best evidence' or some other grading of evidence scheme used.**	For Question 8, if no attempt has been made to combine findings and no statement is made regarding the inappropriateness of combining findings, check "No". if a summary (general) estimate is given anywhere in the abstract, the discussion, or the summary section of the paper, and it is not reported how that estimate was derived, mark "No" even if there is a statement regarding the limitations of combining the findings of the studies reviewed. If in doubt, mark "Can't tell". For an overview to be scored as "Yes" in Question 9, data (not just citations) must be reported that support the main conclusions regarding the primary question(s) that the overview addresses. The score for Question 10, the overall scientific quality, should be based on your answers to the first nine questions. The following guidelines can be used to assist with deriving a summary score: If the "Can't tell" option is used one or more times on the preceding questions, a review is likely to have minor flaws at best and it is difficult to rule out major flaws (a score of 4 or lower). If the "No" option is used on Question 2, 4, 6 or 8, the review is likely to have major flaws (a score of 3 or less, depending on the number and degree of the flaws).
8. Were the findings combined appropriately? *Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?* **"Yes" if the review performs a test for heterogeneity before pooling, does appropriate subgroup testing, appropriate sensitivity analysis, or other such analysis.**	
9. Were the conclusions supported by the reported data? *Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?*	

Systematic Reviews[3, 4]						
10. What was the overall scientific quality of the overview? *How would you rate the scientific quality of this overview?*						
Each Question is scored as Yes, Partially/Can't tell or No						
Extensive Flaws		**Major Flaws**		**Minor Flaws**		**Minimal Flaws**
1	2	3	4	5	6	7

Controlled Trials

Randomized studies

Assessment of Internal Validity

1. Was the assignment to treatment groups really random?
    - Adequate approaches to sequence generation:
        - Computer-generated random numbers
        - Random numbers tables
    - Inferior approaches to sequence generation:
        - Use of alternation, case record number, date of birth, or day of week
    - Not reported
2. Was the treatment allocation concealed?
    - Adequate approaches to concealment of randomization:
        - Centralized or pharmacy-controlled randomization
        - Serially-numbered identical containers
        - On-site computer based system with a randomization sequence that is not readable until allocation
        - Other approaches sequence to clinicians and patients
    - Inferior approaches to concealment of randomization:
        - Use of alternation, case record number, date of birth, or day of week
        - Open random numbers lists
        - Serially numbered envelopes (even sealed opaque envelopes can be subject to manipulation)
    - Not reported

3. Were the groups similar at baseline in terms of prognostic factors?
4. Were the eligibility criteria specified?
5. Were outcome assessors blinded to the treatment allocation?
6. Was the care provider blinded?
7. Was the patient kept unaware of the treatment received?
8. Did the article include an intention-to-treat analysis or provide the data needed to calculate it (that is, the number assigned to each group, number of subjects who finished in each group, and their results)?
9. Did the study maintain comparable groups?
10. Did the article report attrition, crossovers, adherence, and contamination?

11. Is there important differential loss to follow-up or overall high loss to follow-up? (Give numbers in each group.)

Assessment of External Validity (Generalizability)

1. How similar is the population to the population to whom the intervention would be applied?
2. How many patients were recruited?
3. What were the exclusion criteria for recruitment? (Give numbers excluded at each step.)
4. What was the funding source and role of funder in the study?
5. Did the control group receive the standard of care?
6. What was the length of follow-up? (Give numbers at each stage of attrition.)

Non-randomized studies

Assessment of Internal Validity

1. Was the selection of patients for inclusion nonbiased; that is, was any group of patients systematically excluded?
2. Is there important differential loss to followup or overall high loss to followup? (Give numbers in each group.)
3. Were the investigated events specified and defined?
4. Was there a clear description of the techniques used to identify the events?
5. Was there nonbiased and accurate ascertainment of events (independent ascertainers; validation of ascertainment technique)?
6. Were potential confounding variables and risk factors identified and examined using acceptable statistical techniques?
7. Was the duration of follow-up reasonable timing for investigated events? (Did it meet the stated threshold?)

Assessment of External Validity

1. Was the description of the population adequate?
2. How similar is the population to the population to whom the intervention would be applied?
3. How many patients were recruited?
4. What were the exclusion criteria for recruitment? (Give numbers excluded at each step.)
5. What was the funding source and role of funder in the study?

References for appendix B:

1. Anonymous. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews CRD report number 4 (2nd edition. *CRD.* 2001.

2. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *American Journal of Preventive Medicine.* 2001;20(3 Suppl):21-35.

3. Furlan AD, Clarke J, Rosmin E. A critical review of reviews on the treatment of chronic low back pain. *Spine.* 2001;26(7):E155-E162.
4. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology.* 1991;44(11):1271-1278.

Appendix C. Excluded studies

Reasons for exclusion:
1=Foreign language
2= wrong outcome
3= wrong intervention
4=wrong population
5=wrong publication type
6=wrong study design
7= insufficient duration

Excluded publications	Code
Allen BR, Lakhanpaul M, Morris A, et al. Systemic exposure, tolerability, and efficacy of pimecrolimus cream 1% in atopic dermatitis patients. Archives of Disease in Childhood. Nov 2003;88(11):969-973.	6
Ashcroft D, Chen L-C, Garside R, Stein K, Williams H. Topical pimecrolimus for eczema. Cochrane Database of Systematic Reviews. 2007;4:4.	3
Barba JF, Beirana A, Cohen V, et al. Pimecrolimus cream 1% is effective, well-tolerated and safe in infants and children with atopic eczema of the face. Journal of the European Academy of Dermatology & Venereology. 2003;17.	5
Barbier N, Paul C, Luger T, et al. Validation of the Eczema Area and Severity Index for atopic dermatitis in a cohort of 1550 patients from the pimecrolimus cream 1% randomized controlled clinical trials programme. British Journal of Dermatology. Jan 2004;150(1):96-102.	6
Beck LA. The efficacy and safety of tacrolimus ointment: a clinical review. Journal of the American Academy of Dermatology. Aug 2005;53(2 Suppl 2):S165-170.	6
Berger TG, Duvic M, Van Voorhees AS, VanBeek MJ, Frieden IJ, American Academy of Dermatology Association Task F. The use of topical calcineurin inhibitors in dermatology: safety concerns. Report of the American Academy of Dermatology Association Task Force.[erratum appears in J Am Acad Dermatol. 2006 Aug;55(2):271 Note: VanBeek, Marta J [added]]. Journal of the American Academy of Dermatology. May 2006;54(5):818-823.	6
Bernard LA, Eichenfield LF. Topical immunomodulators for atopic dermatitis. Current Opinion in Pediatrics. Aug 2002;14(4):414-418.	6
Beyerler M, Schmid-Grendelmeier P, Hafner J. Significantly elevated systemic levels after occlusive application of topical tacrolimus in atopic dermatitis. Dermatology. 2006;212(3):260-261.	6
Bos JD. Topical tacrolimus and pimecrolimus are not associated with skin atrophy. British Journal of Dermatology. Feb 2002;146(2):342; author reply 343.	5

Excluded publications	Code
Bos JD. Non-steroidal topical immunomodulators provide skin-selective, self-limiting treatment in atopic dermatitis. European Journal of Dermatology. Sep-Oct 2003;13(5):455-461.	6
Breuer K, Braeutigam M, Kapp A, Werfel T. Influence of pimecrolimus cream 1% on different morphological signs of eczema in infants with atopic dermatitis. Dermatology. 2004;209(4):314-320.	5
Breuer K, Werfel T, Kapp A. Safety and efficacy of topical calcineurin inhibitors in the treatment of childhood atopic dermatitis. American Journal of Clinical Dermatology. 2005;6(2):65-77.	6
Burkhart CG, Burkhart CN. Tacrolimus and topical steroids: which is more effective? British Journal of Dermatology. Dec 2004;151(6):1281; author reply 1281-1282.	5
Burkhart CN, Burkhart CG. Pilot study of patient satisfaction with nonfluorinated topical steroids compared with a topical immunomodulator in atopiform dermatitis. International Journal of Dermatology. Mar 2004;43(3):215-219.	7
Callen J, Chamlin S, Eichenfield LF, et al. A systematic review of the safety of topical therapies for atopic dermatitis. British Journal of Dermatology. Feb 2007;156(2):203-221.	6
Castro APBM. Calcineurin inhibitors in the treatment of allergic dermatitis. Jornal de Pediatria. Nov 2006;82(5 Suppl):S166-172.	6
Cheer SM, Plosker GL. Tacrolimus ointment. A review of its therapeutic potential as a topical therapy in atopic dermatitis. American Journal of Clinical Dermatology. 2001;2(6):389-406.	6
Cho M, Puma I, Nguyen D, Schut R, Glesne L. Development of Kaposi's sarcoma in an AIDS patient after treatment with topical tacrolimus. Journal of the American Academy of Dermatology. Jan 2004;50(1):149-150.	6
Cohen B. Review of pimecrolimus cream 1% in children for the treatment of mild to moderate atopic dermatitis. Clinical Pediatrics. Jan 2007;46(1):7-15.	6
Dahl RE, Bernhisel-Broadbent J, Scanlon-Holdford S, Sampson HA, Lupo M. Sleep disturbances in children with atopic dermatitis. Arch Pediatr Adolesc Med. August 1, 1995 1995;149(8):856-860.	4
Eichenfield LF, Ho V, Matsunaga J, Leclerc P, Paul C, Hanifin JM. Blood concentrations, tolerability and efficacy of pimecrolimus cream 1% in Japanese infants and children with atopic dermatitis. Journal of Dermatology. Apr 2007;34(4):231-236.	6
Fivenson D, Arnold RJ, Kaniecki DJ, Cohen JL, Frech F, Finlay AY. The effect of atopic dermatitis on total burden of illness and quality of life on adults and children in a large managed care organization. J Manag Care Pharm. Sep-Oct 2002;8(5):333-342.	6
Freeman AK, Serle J, VanVeldhuisen P, et al. Tacrolimus ointment in the treatment of eyelid dermatitis. Cutis. Apr 2004;73(4):267-271.	6
Furue M, Terao H, Moroi Y, et al. Dosage and adverse effects of topical tacrolimus and steroids in daily management of atopic dermatitis. Journal of Dermatology. Apr 2004;31(4):277-283.	6

Excluded publications	Code
Gianni LM, Sulli MM. Topical tacrolimus in the treatment of atopic dermatitis. Annals of Pharmacotherapy. Jul-Aug 2001;35(7-8):943-946.	6
Hanifin JM, Paller AS, Eichenfield L, et al. Efficacy and safety of tacrolimus ointment treatment for up to 4 years in patients with atopic dermatitis. J Am Acad Dermatol. Aug 2005;53(2 Suppl 2):S186-194.	6
Harper J, Green A, Scott G, et al. First experience of topical SDZ ASM 981 in children with atopic dermatitis. British Journal of Dermatology. Apr 2001;144(4):781-787.	6
Hebert AA. Review of pimecrolimus cream 1% for the treatment of mild to moderate atopic dermatitis. Clinical Therapeutics. Dec 2006;28(12):1972-1982.	6
Hidalgo BP, Knight T, Burls A. A systematic review of effectiveness and cost effectiveness of tacrolimus ointment for topical treatment of atopic dermatitis in adults and children. Database of Abstracts of Reviews of Effects. 2007(4).	3
Hoare C, Li Wan Po A, Williams H. Systematic review of treatments for atopic eczema. Health Technol Assess. 2000;4(37):1-91.	3
Hon K-LE, Lam M-CA, Leung T-F, Chow C-M, Wong E, Leung AKC. Assessing itch in children with atopic dermatitis treated with tacrolimus: objective versus subjective assessment. Advances in Therapy. Jan-Feb 2007;24(1):23-28.	7
Kang S. Tacrolimus ointment for adults with moderate to severe atopic dermatitis: a dose escalation study (abstract 1253). J Invest Dermatol. 1998;110:681.	5
Kang S, Lucky AW, Pariser D, Lawrence I, Hanifin JM. Long-term safety and efficacy of tacrolimus ointment for the treatment of atopic dermatitis in children. Journal of the American Academy of Dermatology. Jan 2001;44(1 Suppl):S58-64.	6
Katoh N, Hirano S, Yasuno H, Kishimoto S. Effects of tacrolimus ointment on facial eruption, itch, and scratching in patients with atopic dermatitis. Journal of Dermatology. Mar 2004;31(3):194-199.	6
Kawakami T, Soma Y, Morita E, et al. Safe and effective treatment of refractory facial lesions in atopic dermatitis using topical tacrolimus following corticosteroid discontinuation. Dermatology. 2001;203(1):32-37.	5
Kiebert G, Sorensen SV, Revicki D, et al. Atopic dermatitis is associated with a decrement in health-related quality of life. International Journal of Dermatology. 2002;41(3):151-158.	6
Koo JY, Fleischer AB, Jr., Abramovits W, et al. Tacrolimus ointment is safe and effective in the treatment of atopic dermatitis: results in 8000 patients. J Am Acad Dermatol. Aug 2005;53(2 Suppl 2):S195-205.	6
Korfitis C, Gregoriou S, Rallis E, Rigopoulos D. Pimecrolimus versus topical corticosteroids in dermatology. Expert Opinion on Pharmacotherapy. Jul 2007;8(10):1565-1573.	6
Kreuter A, Hochdorfer B, Altmeyer P, Gambichler T. Pimecrolimus 1% cream for perianal atopic dermatitis. British Journal of Dermatology. Jan 2005;152(1):186-187.	6

Excluded publications	Code
Kyllonen H, Remitz A, Mandelin JM, Elg P, Reitamo S. Effects of 1-year intermittent treatment with topical tacrolimus monotherapy on skin collagen synthesis in patients with atopic dermatitis. British Journal of Dermatology. Jun 2004;150(6):1174-1181.	2
Lan C-CE, Huang C-C, Chen Y-T, Wang L-F, Lin C-T, Chen G-S. Tacrolimus ointment for the treatment of atopic dermatitis: report of first clinical experience in Taiwan. Kaohsiung Journal of Medical Sciences. Jun 2003;19(6):296-304.	6
Langeland T, Engh V. Topical use of tacrolimus and squamous cell carcinoma on the penis. British Journal of Dermatology. Jan 2005;152(1):183-185.	6
Lawson V, Lewis-Jones MS, Finlay AY, Reid P, Owens RG. The family impact of childhood atopic dermatitis: the Dermatitis Family Impact Questionnaire. Br J Dermatol. Jan 1998;138(1):107-113.	4
Linnet J, Jemec GB. An assessment of anxiety and dermatology life quality in patients with atopic dermatitis. Br J Dermatol. Feb 1999;140(2):268-272.	3
Lonsdale-Eccles AA, Velangi S. Herpes simplex of the vulva evoked by topical tacrolimus treatment. Clinical & Experimental Dermatology. Jan 2005;30(1):95-96.	6
Lubbe J, Friedlander SF, Cribier B, et al. Safety, efficacy, and dosage of 1% pimecrolimus cream for the treatment of atopic dermatitis in daily practice. American Journal of Clinical Dermatology. 2006;7(2):121-131.	6
Milingou M, Antille C, Sorg O, Saurat J-H, Lubbe J. Alcohol intolerance and facial flushing in patients treated with topical tacrolimus. Archives of Dermatology. Dec 2004;140(12):1542-1544.	6
Munzenberger PJ, Montejo JM. Safety of topical calcineurin inhibitors for the treatment of atopic dermatitis. Pharmacotherapy. Jul 2007;27(7):1020-1028.	6
Nakagawa H. Comparative study of FK506 (tacrolimus) ointment vs alclometasone dipropionate ointment in atopic dermatitis (face and neck lesions) (abstract 1266). J Invest Dermatol. 1998;110:683.	5
Nakagawa H. Comparison of the efficacy and safety of 0.1% tacrolimus ointment with topical corticosteroids in adult patients with atopic dermatitis: review of randomised, double-blind clinical studies conducted in Japan. Clinical Drug Investigation. 2006;26(5):235-246.	6
Nakahara T, Koga T, Fukagawa S, Uchi H, Furue M. Intermittent topical corticosteroid/tacrolimus sequential therapy improves lichenification and chronic papules more efficiently than intermittent topical corticosteroid/emollient sequential therapy in patients with atopic dermatitis. Journal of Dermatology. Jul 2004;31(7):524-528.	6
Naylor M, Elmets C, Jaracz E, Rico JM. Non-melanoma skin cancer in patients with atopic dermatitis treated with topical tacrolimus. Journal of Dermatological Treatment. Aug 2005;16(3):149-153.	6
Pacor ML, Di Lorenzo G, Martinelli N, Mansueto P, Rini GB, Corrocher R. Comparing tacrolimus ointment and oral cyclosporine in adult patients affected by atopic dermatitis: a randomized study. Clinical & Experimental Allergy. Apr 2004;34(4):639-645.	3

Excluded publications	Code
Papp KA, Werfel T, Folster-Holst R, et al. Long-term control of atopic dermatitis with pimecrolimus cream 1% in infants and young children: a two-year study. Journal of the American Academy of Dermatology. Feb 2005;52(2):240-246.	6
Papp K, Staab D, Harper J, et al. Effect of pimecrolimus cream 1% on the long-term course of pediatric atopic dermatitis. International Journal of Dermatology. Dec 2004;43(12):978-983.	6
Park CW, Lee BH, Lee CH. Tacrolimus reduces staphylococcal colonization on the skin in Korean atopic dermatitis patients. Drugs Under Experimental & Clinical Research. 2005;31(2):77-87.	2
Patel RR, Vander Straten MR, Korman NJ. The safety and efficacy of tacrolimus therapy in patients younger than 2 years with atopic dermatitis. Archives of Dermatology. Sep 2003;139(9):1184-1186.	6
Patel TS, Greer SC, Skinner RB, Jr. Cancer concerns with topical immunomodulators in atopic dermatitis: overview of data and recommendations to clinicians. American Journal of Clinical Dermatology. 2007;8(4):189-194.	6
Paul C, Cork M, Rossi AB, Papp KA, Barbier N, de Prost Y. Safety and tolerability of 1% pimecrolimus cream among infants: experience with 1133 patients treated for up to 2 years. Pediatrics. Jan 2006;117(1):e118-128.	6
Reitamo S. 0.1% tacrolimus ointment is an effective treatment for adults with moderate or severe atopic dermatitis. J Eur Acad Dermatol Cenereol. 2003;17:20-31.	6
Remitz A, Harper J, Rustin M, et al. Long-term safety and efficacy of tacrolimus ointment for the treatment of atopic dermatitis in children. Acta Dermato-Venereologica. 2007;87(1):54-61.	6
Ricci G, Dondi A, Patrizi A. Role of topical calcineurin inhibitors on atopic dermatitis of children. Current Medicinal Chemistry. 2007;14(14):1579-1591.	6
Rikkers SM, Holland GN, Drayton GE, Michel FK, Torres MF, Takahashi S. Topical tacrolimus treatment of atopic eyelid disease. American Journal of Ophthalmology. Mar 2003;135(3):297-302.	6
Rodriguez-Martin M, Saez-Rodriguez M, Carnerero-Rodriguez A, et al. Treatment of perioral dermatitis with topical pimecrolimus. Journal of the American Academy of Dermatology. Mar 2007;56(3):529-530.	6
Ruiz-Maldonado R. Pimecrolimus related crusted scabies in an infant. Pediatric Dermatology. May-Jun 2006;23(3):299-300.	6
Ruzicka T, Assmann T, Homey B. Tacrolimus: the drug for the turn of the millennium? Archives of Dermatology. May 1999;135(5):574-580.	6
Salavec M, Buckova H. [First experiences with 1% pimecrolimus cream therapy in prevention of atopic eczema flares in children]. Ceskoslovenska Dermatologie. 2004;79(1):3-7.	1
Sand C. Topical tacrolimus ointment may induce skin tags in treated patients. Acta Dermato-Venereologica. 2003;83(4):317.	6

Excluded publications	Code
Schiffner R, Schiffner-Rohe J, Landthaler M, Stolz W. Treatment of atopic dermatitis and impact on quality of life: a review with emphasis on topical non-corticosteroids. Pharmacoeconomics. 2003;21(3):159-179.	6
Schurmeyer-Horst F, Luger TA, Bohm M. Long-term efficacy of occlusive therapy with topical pimecrolimus in severe dyshidrosiform hand and foot eczema. Dermatology. 2007;214(1):99-100.	6
Segura S, Romero D, Carrera C, Iranzo P, Estrach T. Eczema herpeticum during treatment of atopic dermatitis with 1% pimecrolimus cream. Acta Dermato-Venereologica. 2005;85(6):524-525.	6
Shainhouse T, Eichenfield LF. Long-term safety of tacrolimus ointment in children treated for atopic dermatitis. Expert Opinion on Drug Safety. Sep 2003;2(5):457-465.	6
Simon D, Lubbe J, Wuthrich B, et al. Benefits from the use of a pimecrolimus-based treatment in the management of atopic dermatitis in clinical practice. Analysis of a Swiss cohort. Dermatology. 2006;213(4):313-318.	6
Simpson D, Noble S. Tacrolimus ointment: a review of its use in atopic dermatitis and its clinical potential in other inflammatory skin conditions. Drugs. 2005;65(6):827-858.	6
Singalavanija S, Noppakun N, Limpongsanuruk W, et al. Efficacy and safety of tacrolimus ointment in pediatric Patients with moderate to severe atopic dermatitis. Journal of the Medical Association of Thailand. Nov 2006;89(11):1915-1922.	6
Skaehill PA. Tacrolimus in dermatologic disorders. Annals of Pharmacotherapy. May 2001;35(5):582-588.	6
Staab D, Pariser D, Gottlieb AB, et al. Low systemic absorption and good tolerability of pimecrolimus, administered as 1% cream (Elidel) in infants with atopic dermatitis--a multicenter, 3-week, open-label study. Pediatric Dermatology. Sep-Oct 2005;22(5):465-471.	6
Tan J, Langley R. Safety and efficacy of tacrolimus ointment 0.1% (Protopic) in atopic dermatitis: a Canadian open-label multicenter study. Journal of Cutaneous Medicine & Surgery. Jul-Aug 2004;8(4):213-219.	6
Thaci D, Steinmeyer K, Ebelin M-E, Scott G, Kaufmann R. Occlusive treatment of chronic hand dermatitis with pimecrolimus cream 1% results in low systemic exposure, is well tolerated, safe, and effective. An open study. Dermatology. 2003;207(1):37-42.	6
Thelmo MC, Lang W, Brooke E, et al. An open-label pilot study to evaluate the safety and efficacy of topically applied tacrolimus ointment for the treatment of hand and/or foot eczema. Journal of Dermatological Treatment. Sep 2003;14(3):136-140.	6
Thestrup-Pedersen K. Tacrolimus treatment of atopic eczema/dermatitis syndrome. Current Opinion in Allergy & Clinical Immunology. Oct 2003;3(5):359-362.	6
Virtanen HM, Reitamo S, Kari M, Kari O. Effect of 0.03% tacrolimus ointment on conjunctival cytology in patients with severe atopic blepharoconjunctivitis: a retrospective study. Acta Ophthalmologica Scandinavica. Oct 2006;84(5):693-695.	6

Excluded publications	Code
Weinberg JM. Formulary review of therapeutic alternatives for atopic dermatitis: focus on pimecrolimus. Journal of Managed Care Pharmacy. Jan-Feb 2005;11(1):56-64.	6
Weinberg JM, Bowerman JG, Brown SM, et al. Atopic dermatitis: a new treatment paradigm using pimecrolimus. Journal of Drugs in Dermatology: JDD. Apr 2003;2(2):131-140.	6
Wellington K, Jarvis B. Spotlight on topical pimecrolimus in atopic dermatitis. American Journal of Clinical Dermatology. 2002;3(6):435-438.	6
Wellington K, Noble S. Pimecrolimus: a review of its use in atopic dermatitis. American Journal of Clinical Dermatology. 2004;5(6):479-495.	6
Yeung CK, Ma KC, Chan HHL. Efficacy and safety of tacrolimus ointment monotherapy in chinese children with atopic dermatitis. SKINmed. Jan-Feb 2006;5(1):12-17.	6

Appendix D. Glossary

Following is a listing of terms commonly used in reports produced by the Drug Effectiveness Review Project *as they apply to these reports*. For that reason, some terms definitions may vary slightly from other published definitions.

Adherence: Following the course of treatment proscribed by a study protocol.

Adverse effect: An *adverse event* for which the causal relation between the drug/intervention and the event is at least a reasonable possibility.

Adverse event: An adverse outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it.

Active-control trial: A trial comparing a drug in a particular class or group to another drug outside of that class or group.

Allocation concealment: The process by which the person determining randomization is blinded to a study participant's group allocation.

Before-after study: A type of nonrandomized study in which data are collected before and after patients receive an intervention. Before-after studies can have a single arm or can include a control group.

Bias: A systematic error or deviation in results or inferences from the truth. Several types of bias can appear in published trials, including selection bias, performance bias, detection bias, and reporting bias.

Blinding: The process of preventing those involved in a trial from knowing which comparison group a particular participant belongs to. Trials are frequently referred to as "double-blind" without further explanation of whether this description refers to patients, caregivers, investigators, or other study staff.

Case series: A study reporting observations on a series of patients all receiving the same intervention (no control group).

Case study: A study reporting observations on a single patient.

Case-control study: A study that compares people with a specific disease or outcome of interest (cases) to people from the same population without that disease or outcome (controls.)

Clinically significant: A result that is large enough to affect a patient's disease state in a manner that is noticeable to the patient and/or caregiver.

Cohort study: An observational study in which a defined group of people (the cohort) is followed over time and compared with a group of people who were exposed or not exposed to a particular intervention or other factor of interest. A prospective cohort study assembles participants and follows them into the future. A retrospective cohort study identifies subjects from past records and follows them from the time of those records to the present.

Confidence interval: The range of values calculated from the data such that there is a level of confidence, or certainty, that it contains the true value. A 95% confidence interval is generally used in Drug Effectiveness Review Project reports.

Confounder: A factor that is associated with both an intervention and an outcome of interest.

Controlled clinical trial: A clinical trial that includes a control group but no or inadequate methods of randomization.

Convenience sample: A group of individuals being studied because they are conveniently accessible in some way. Convenience samples may or may not be representative of a population that would normally be receiving an intervention.

Crossover trial: A type of clinical trial comparing two or more interventions in which the participants complete the course of one intervention and then are switched to another intervention.

Direct analysis: The practice of using data from head-to-head trials to draw conclusions about the comparative effectiveness of drugs within a class or group. Results of direct analysis are the preferred source of data in Drug Effectiveness Review Project reports.

Dose-response relationship: The relationship between the quantity of treatment given and its effect on outcome. In meta-analysis, dose-response relationships can be investigated using meta-regression.

Double-blind: Those involved in a trial have been prevented from knowing which comparison group a particular participant belongs to. Double-blind trials can include blinding of patients, caregivers, investigators, and/or other study staff.

Double-dummy: The use of two placebos that mimic the appearance and administration of a trial's active interventions when those interventions differ in appearance or method of administration (for example, an oral agent compared with an injectable agent.)

Effectiveness: The extent to which a specific intervention used under ordinary circumstances does what it is intended to do.

Effectiveness outcomes: Outcomes that are generally important to patients and caregivers, such as quality of life, hospitalizations, and ability to work. Data on effectiveness outcomes usually comes from longer-term studies of a "real-world" population.

Efficacy: The extent to which an intervention produces a beneficial result under ideal conditions in a selected and controlled population.

Estimate of effect: The observed relationship between an intervention and an outcome. Estimate of effect can be expressed in a number of ways, including number needed to treat, odds ratio, risk difference, and risk ratio.

Equivalence trial: A trial designed to determine whether the response to two or more treatments differs by an amount that is clinically unimportant. This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence level of clinically acceptable differences.

External validity: The extent to which reported results are generalizable to a relevant population.

Fixed-effect model: A model that calculates a pooled effect estimate using the assumption that all observed variation between studies is caused by chance. Studies are assumed to be measuring the same overall effect. An alternative model is the random-effects model.

Forest plot: A graphical representation of the individual results of each study included in a meta-analysis together with the combined meta-analysis result. The plot also allows readers to see the heterogeneity among the results of the studies. The results of individual studies are shown as squares centered on each study's point estimate. A horizontal line runs through each square to show each study's confidence interval—usually a 95% confidence interval. The overall estimate from the meta-analysis and its confidence interval are shown as a diamond at the bottom of the plot. The center of the diamond represents the pooled point estimate, and its horizontal tips represent the confidence interval.

Funnel plot: A graphical display of some measure of study precision plotted against effect size that can be used to determine whether a link exists between study size and treatment effect.

Generalizability: see *External Validity*

Hazard ratio: The increased risk with which one group is likely to experience an outcome of interest. It is similar to a risk ratio. For example, if the hazard ratio for death for a treatment is 0.5, then we can say that treated patients are likely to die at half the rate of untreated patients.

Head-to-head trial: A trial that directly compares one drug of a particular class or group to another in the same class or group.

Heterogeneity: The variation in or diversity of participants, interventions, and measurement of outcomes across a set of studies.

Indirect analysis: The practice of using data from trials comparing one drug in a particular class or group to a drug outside that class or group or to placebo and attempting to draw conclusions about the comparative effectiveness of drugs within a class or group based on those data. For

example, using direct comparisons between drugs A and B and between drugs B and C to make indirect comparisons between drugs A and C.

Intention to treat (ITT): The use of data from a randomized controlled trial in which data from all randomized patients are accounted for in the final results. Trials often report results as being based on intention to treat despite some patients being excluded from the analysis.

Internal validity: The extent to which the design and conduct of a study are likely to have prevented bias. Generally, the higher the interval validity, the better the quality of the published study.

Inter-rater reliability: The degree of stability exhibited when a measurement is repeated under identical conditions by different raters.

Intermediate outcome: An outcome not of direct practical importance but believed to reflect outcomes that are important. For example, blood pressure is not directly important to patients but it is often used as an outcome in clinical trials because it is a risk factor for stroke and myocardial infarction.

Logistic regression: A form of regression analysis that models an individual's odds of disease or other outcome as a function of a risk factor or intervention.

Mean difference: A method used to combine measures on continuous scales (such as weight), where the mean, standard deviation, and sample size in each group are known.

Meta-analysis: The use of statistical techniques in a systematic review to integrate the results of included studies. Although they are sometimes used interchangeably, meta-analyses are not synonymous with systematic reviews. However, systematic reviews often include meta-analyses.

Meta-regression: A technique used to explore the relationship between study characteristics (for example, concealment of allocation, baseline risk factors, or timing of the intervention) and study results (for example, the magnitude of effect observed in each study) in a systematic review.

Multivariate analysis: Measuring the impact of more than one variable at a time while analyzing a set of data.

N of 1 trial: A randomized trial in an individual to determine the optimum treatment for that individual.

Noninferiority trial: A trial designed to determine whether the effect of a new treatment is not worse than a standard treatment by more than a prespecified amount. A one-sided version of an equivalence trial.

Nonrandomized study: Any study estimating the effectiveness of an intervention (harm or benefit) that does not use randomization to allocate patients to comparison groups. There are

many types of nonrandomized studies, including cohort studies, case-control studies, and before -after studies.

Null hypothesis: The statistical hypothesis that one variable (for example, which treatment a study participant was allocated to receive) has no association with another variable or set of variables.

Number needed to treat (NNT): An estimate of how many people need to receive a treatment before one person would experience a beneficial outcome.

Observational study: A type of nonrandomized study in which the investigators do not intervene, instead simply observing the course of events.

Odds ratio (OR): The ratio of the odds of an event in one group to the odds of an event in another group. An odds ratio of 1.0 indicates no difference between comparison groups. For undesirable outcomes an OR that is <1.0 indicates that the intervention was effective in reducing the risk of that outcome.

One-tailed test: A hypothesis test in which the values for which we can reject the null hypothesis are located entirely in one tail of the probability distribution. For example, testing whether one treatment is better than another (rather than testing whether one treatment is either better or worse than another).

Open-label trial: A clinical trial in which the investigator and participant are aware which intervention is being used for which participant (that is, a trial that is not blinded). Random allocation may or may not be used in open-label trials.

Per protocol: The subset of participants from a randomized controlled trial who complied with the protocol sufficiently to ensure that their data would be likely to exhibit the effect of treatment. Per protocol analyses are sometimes misidentified in published trials as intention-to-treat analyses.

Point estimate: The results (for example, mean, weighted mean difference, odds ratio, risk ratio, or risk difference) obtained in a sample (a study or a meta-analysis) that are used as the best estimate of what is true for the relevant population from which the sample is taken.

Pooling: The practice of combining data from several studies to draw conclusions about treatment effects.

Power: The probability that a trial will detect statistically significant differences among intervention effects. Studies with small sample sizes can frequently be insufficiently powered to detect difference.

Precision: The likelihood of random errors in the results of a study, meta-analysis, or measurement. The greater the precision, the less random error. Confidence intervals around the

estimate of effect from each study are one way of expressing precision, with a narrower confidence interval meaning more precision.

Prospective study: A study in which people are identified according to current risk status or exposure and followed forward through time to observe outcome.

Publication bias: A bias caused by availability of only a subset of all relevant data. The publication of research can depend on the nature and direction of study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (for example, only outcomes or subgroups in which a statistically significant difference was found).

P value: The probability (ranging from zero to one) that the results observed in a study could have occurred by chance if the null hypothesis was true. A *P* value of ≤0.05 is often used as a threshold to indicate statistical significance.

Random-effects model: A statistical model in which both within-study sampling error (variance) and between-studies variation are included in the assessment of the uncertainty (confidence interval) of the results of a meta-analysis. When there is heterogeneity among the results of the included studies beyond chance, random-effects models will give wider confidence intervals than fixed-effect models.

Randomization: A process by which study participants are allocated to treatment groups in a trial. Adequate (that is, unbiased) methods of randomization include computer generated schedules and random numbers tables.

Randomized controlled trial (RCT): A trial in which two or more interventions are compared through random allocation of participants.

Regression analysis: A statistical modeling technique used to estimate or predict the influence of one or more independent variables on a dependent variable, e.g. for example, the effect of age, sex, and confounding disease on the effectiveness of an intervention.

Relative risk (RR): The ratio of risks in two groups; same as a risk ratio.

Retrospective study: A study in which the outcomes have occurred prior to study entry.

Risk difference: The difference in size of risk between two groups.

Risk ratio (RR): The ratio of risks in two groups. In intervention studies, it is the ratio of the risk in the intervention group to the risk in the control group. A risk ratio of 1 indicates no difference between comparison groups. For undesirable outcomes, a risk ratio that is <1 indicates that the intervention was effective in reducing the risk of that outcome. Same as relative risk.

Sensitivity analysis: An analysis used to determine to what extent the results of a study or systematic review are sensitive to changes in how the study was done. Sensitivity analyses are used to assess how robust the results are to uncertain decisions or assumptions about the data and the methods that were used.

Standard deviation (SD): A measure of the spread or dispersion of a set of observations, calculated as the average difference from the mean value in the sample.

Standard error (SE): A measure of the variation in the sample statistic over all possible samples of the same size. The standard error decreases as the sample size increases.

Statistically significant (SS): A result that is unlikely to have happened by chance.

Subgroup analysis: An analysis in which an intervention is evaluated in a defined subset of the participants in a trial, such as by sex or in age categories.

Superiority trial: A trial designed to test whether one intervention is superior to another.

Systematic review: A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research and to collect and analyze data from the studies that are included in the review.

Tolerability: Ability of a drug to be tolerated by the patient. Tolerability is affected by adverse effects that can affect quality of life and willingness to continue treatment, although they are usually transient and not clinically significant.

Type I error: A conclusion that there is evidence that a treatment works when it actually does not work (false positive).

Type II error: A conclusion that there is no evidence that a treatment works when it actually does work (false-negative).

Validity: The degree to which a result (of a measurement or study) is likely to be true and free of bias (systematic errors).

Appendix E. Results for sensitivity analysis (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)

Trial	Mean patient age	Tacrolimus 0.03% (n=158)	Pimecrolimus 1% (n=390)	Vehicle (n=358)
Schachner 2005[37]	6.9 years	50.6%	N/A	25.8%
Ho 2003[36]	13 months	N/A	54.5%	23.8%
Eichenfield 2002[22] study #305	6.7 years	N/A	37.7%	16.2%
Eichenfield 2002[22] study #307	6.8 years	N/A	32.1%	20.6%
Pooled rate		50.6%	41.3%	22.1%
95% CI		(42.8% to 58.4%)	(28.4% to 54.2%)	(17.8% to 26.5%)
Heterogeneity statistics	Cochrane's Q	N/A	14.5 (*P*<0.001)	3.1 (*P*=0.38)
	I^2	N/A	86.2%	2.8%
Indirect pooled relative risk (95% CI): 0.97 (0.63–1.48)				

Excluding Ho, 2003[a]

Pooled rate		34.7%	21.4%
95% CI		(29.0% to 40.4%)	(15.6% to 27.3%)
Heterogeneity statistics	Cochrane's Q	1.0 (*P*=0.32)	2.97 (*P*=0.23)
	I^2	0%	32.6%
Indirect pooled relative risk (95% CI): 1.05 (0.64 to 1.72)			

[a] sensitivity analysis excluding Ho, 2003 were conducted to determine whether differences in age in this trial (among pimecrolimus trials) would significantly affect the pooled results.

Appendix F. Assessing the strength of comparative evidence (using the modified GRADE approach)

We assessed the overall strength of evidence for the efficacy/effectiveness of included trials. The overall strength of evidence for a particular key question reflects the design, quality, consistency, directness, and precision of the effect estimate**.** We rate the overall strength of evidence as low, moderate, high, or insufficient using a modified GRADE approach established by the Evidence-based Practice Centers. *High* strength of evidence indicates high confidence in the estimate of effect and that the evidence reflects the true effect; further research is unlikely to change our confidence. *Moderate* strength of evidence indicates moderate confidence that the evidence reflects the true effect; further research may change our confidence in the estimate and may change the estimate. *Low* strength of evidence indicates low confidence that the evidence reflects the true effect; further research is likely to change our confidence in the estimate and is likely to change the estimate. *Insufficient* indicates that evidence is unavailable or does not permit estimation of an effect.

For this review we used a "point system" for each domain to help guide our overall assessment. Listed below is a description of our "point system." Further information regarding the modified GRADE approach can be found in the Effective Health Care Program Methods Manual, http://effectivehealthcare.ahrq.gov.

Risk of bias
(+1): If adequate description of methods for randomization, allocation concealment, blinding, ITT, reasonable withdrawal/drop out rates—would be rated good.
(-1): If limited description of methods for randomization, allocation concealment, and blinding—would be rated fair.
(-2 or -3): Additional points would be taken if there were problems with analysis (for example inadequate ITT population), high withdrawal/drop out rates, and/or significant selective reporting (of outcomes or harms) in addition to inadequate randomization, allocation concealment, and blinding—would be rated poor.

Study design
(+1): If study designs were randomized trials.
(-1): If patients were nonrandomized, open-label, or post-hoc subgroup analyses.

Consistency
(+1): Yes- direction of effect similar across studies.
(-1): No- significant variation in effect across studies.

Directness
(+1): If patient relevant health outcome were measured/evaluated.
(-1): If outcome was an intermediate outcome with some validity demonstrating correlation to main health outcome of interest.
(-2 or -3): Additional points would be taken if outcome was an intermediate outcome with unclear or no validity established and lacking strong correlation to main health outcome of interest or if indirect bodies of evidence were used to make comparisons between interventions.

Precision
(+1): Yes-estimate would allow clinically useful conclusion.
(-1): No-confidence interval is wide enough to include clinically distinct conclusions.

A. Mild to moderate disease

No. of studies (# subjects)	Risk of bias: (study design/ quality)	Consistency	Directness	Precision	Comments	Overall strength of evidence
Outcome: Achieving treatment success						
Tacrolimus 0.03% ointment compared with pimecrolimus 1% cream						
2 (562)	H2H/Fair	+1	-1	+1	Pooled RR, 1.19, 0.98-1.45	Moderate-high
4 (425)	RCT/Fair	+1	-2	+1	Pooled RR, 0.97, 0.63-1.48	Moderate
Tacrolimus 0.1% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
No studies	RCT/Fair	-----	-----	-----	-----	Insufficient
Outcome: Reduction in pruritus						
Tacrolimus 0.03% ointment compared with Pimecrolimus 1% cream						
2 (562)	H2H/Fair	+1	-1	Unable to determine	Did not pool results due to heterogenous outcome reporting but qualitative assessment shows minimal difference	Moderate
4 (425)	RCT/Fair	+1	-2	Unable to determine	Did not pool results due to heterogenous outcome reporting	Unable to make comparison
Tacrolimus 0.1% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
No studies	RCT/Fair	-----	-----	-----	-----	Insufficient
Outcome: Patient assessment of overall disease control						
Tacrolimus 0.03% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
No studies	RCT/Fair	-----	-----	-----	-----	Insufficient
Tacrolimus 0.1% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
No studies	RCT/Fair	-----	-----	-----	-----	Insufficient

Abbreviations: H2H, head-to-head trial; RCT, randomized controlled trial; RR, relative risk; WMD, weighted mean difference

B. Moderate to severe disease

No. of studies (# subjects)	Risk of bias: (study design/ quality)	Consistency	Directness	Precision	Comments	Overall strength of evidence
Outcome: Achieving treatment success						
Tacrolimus 0.03% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
4 (457)	RCT/Fair	+1	-2	-1	Pooled RR, 0.89 (95% CI, 0.38-2.07)	Low
Tacrolimus 0.1% ointment compared with Pimecrolimus 1% cream						
1 (224)	H2H/Fair	Unable to determine (1 study)	-1	+1	RR, 1.83 (95% CI, 1.13-2.96)	Moderate
4 (456)	RCT/Fair	+1	-2	-1	Pooled RR, 1.12 (95% CI, 0.48-2.57)	Low
Outcome: Reduction in pruritus						
Tacrolimus 0.03% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
4 (457)	RCT/Fair	+1	-2	-1	Pooled WMD, 0.86 (95% CI, -0.69 to 2.41)	Low
Tacrolimus 0.1% ointment compared with Pimecrolimus 1% cream						
1 (224)	H2H/Fair	Unable to determine	-1	+1	Change in pruritus score from baseline -3.7 cm compared with -2.0 cm, $P \leq 0.01$	Moderate-low
4 (456)	RCT/Fair	+1	-2	-1	Pooled WMD, 0.74 (95% CI, -0.83 to 2.31)	Low
Outcome: Patient assessment of overall disease control						
Tacrolimus 0.03% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
4 (457)	RCT/Fair	+1	-2	-1	Pooled RR, 0.98 (95% CI, 0.56-1.73)	Low
Tacrolimus 0.1% ointment compared with pimecrolimus 1% cream						
No studies	H2H/Fair	-----	-----	-----	-----	Insufficient
4 (456)	RCT/Fair	+1	-2	-1	Pooled RR, 1.07 (95% CI, 0.53-2.13)	Low

Abbreviations: H2H, head-to-head trial; RCT, randomized controlled trial; RR, relative risk; WMD, weighted mean difference