TITLE

Cochrane reviews and dermatological trials outcome concordance; why Core Outcome Sets could make trial results more usable.

AUTHORS

Jochen Schmitt[1], Toni Lange[1], Jan Kottner[2], Cecilia A. C.Prinsen[3], Tobias Weberschock[4a,b], Elisabeth Hahnel[2], Christian Apfelbacher[5], Susanne Brandstetter[5], Andreas Dreher[4a], Giles Stevens[1], Esther Burden-Teh[6], Natasha Rogers[6], Phyllis Spuls[7], Matthew J Grainge[8], Hywel C. Williams[6], Lena Jacobi[1] on behalf of the Cochrane Skin Core Outcome Set Initiative (CS-COUSIN)

[1] Center for Evidence-based Healthcare, Medical Faculty Carl Gustav Carus, TU Dresden, Germany

[2] Charité-Universitätsmedizin Berlin, Department of Dermatology and Allergy, Germany

[3] Amsterdam UMC, VU University medical center, Department of Epidemiology and Biostatistics, Amsterdam Public Health research institute, Amsterdam, the Netherlands

[4a] Department of Dermatology, Venereology and Allergology, University Hospital Frankfurt, Frankfurt/Main, Germany.

[4b] Working group Evidence-based Medicine Frankfurt, Institute for General Practice, Goethe University Frankfurt, Frankfurt/Main, Germany.

[5] Medical Sociology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Germany

[6] Centre for Evidence-based Dermatology, University of Nottingham, Nottingham, UK

[7] Amsterdam UMC, University of Amsterdam, Department of Dermatology, Amsterdam Public health, Infection and Immunity, Meibergdreef 9, Amsterdam, The Netherlands

[8] Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Nottingham, UK.

Corresponding author:

Jochen Schmitt

Center for Evidence-based Healthcare,

University Hospital and Medical Faculty Carl Gustav Carus, TU Dresden, Germany Haus 12, 3. OG, Zimmer 409 Fetscherstr. 74

01307 Dresden

Jochen.Schmitt@uniklinikum-dresden.de

Word count: Abstact 200 words; main text 3377 words; 1 table; 5 figures; 6 supplementary e-tables

ABSTRACT

Evidence-based healthcare requires that relevant outcomes for patients are included in clinical trials ~~which~~ investigat~~e~~ing treatment effects ~~so that~~allowing subsequent systematic reviews ~~can~~to summarize all relevant evidence to guide clinical practice. Currently, no gold standard of outcome choice for dermatology trials and reviews exists. We systematically assessed ~~the degree of~~ concordance between efficacy outcomes in ~~systematic reviews and their included trials in~~ a random sample of 10 Cochrane Skin systematic reviews~~,~~ and the ~~containing~~ 220 dermatology trials included. Reviews did not include 742 (68%) of the 1,086 trial outcomes. Of the 60 outcomes the reviews sought, 17 (28%) of these were not reported in any trial whilst 12 were assessed in less than 50% of trials. For 11/23 (48%) primary review outcomes meta-analysis was impossible, because trial outcomes were absent or unclear. This small overlap of review/trial outcomes could suggest that trials are not measuring the outcomes perceived to be the most important by patients, clinicians, systematic reviewers and trialists. The lack of standardized outcome measures, poor reporting of outcomes in trials and low concordance of outcomes between reviews and primary studies could be improved by the development and implementation of Core Outcome Sets (COS). These are an agreed minimum set of key outcomes, for specified conditions, to be reported in all trials.

Keywords

Systematic review, trial, outcome, core outcome set, efficacy, meta-analysis, dermatology

INTRODUCTION

In evidence-based healthcare, a key determining factor of the scientific value of clinical trials is the choice of outcome. It is of the utmost importance for trials to choose outcomes that are considered relevant to patients, clinicians and other healthcare professionals as well as decision-makers such as commissioners and policy-makers. Outcome measurement instruments must be reliable, valid, and feasible (Boers, 2014, Boers et al., 1998). Trials using inappropriate outcomes may overestimate, underestimate or overlook the effect of the intervention under investigation (Sinha et al., 2008) and standardization is crucial in order to allow cross-trial comparisons in systematic reviews. Similarly, meta-analyses are only possible with comparable outcomes. Moreover, using different outcomes across clinical trials can hinder clinicians who consider trial data, systematic reviews or meta-analyses in their clinical decisions or when formulating clinical practice guidelines. Low or unclear reliability, validity, and feasibility of many outcome measurement instruments and lack of standardization in outcome assessment in trials have been identified as significant barriers towards evidence-based decision-making (Chalmers and Glasziou, 2009).

In the current absence of formal standardization of outcome reporting in clinical trials, for almost all dermatological diseases, systematic reviewers need to have an overview of the outcomes used in relevant trials before they finally choose review outcomes. Nonetheless, through the involvement of clinical and methodological experts and consumers, the choice of outcomes in Cochrane reviews generally tend to capture the key benefits and harms of treatments. A low degree of overlap between trial and review outcomes would potentially indicate that trials were not measuring the outcomes believed by the wider healthcare community to be particularly important. It would also raise the possibility of weaknesses in systematic review conclusions.

A preliminary overview of the choice of outcomes in all 69 reviews of the Cochrane Skin Group (CSG) published until January 2015 suggested a substantial mismatch between the outcomes of interest to the review and those reported in the included clinical trials (Schmitt et al., 2016). Only 67% (n=271) of 402 predefined review outcomes were found in at least one component trial of the reviews. There are almost no mutually agreed standards for outcome selection in dermatological research. Such agreements aim to make trials relevant and comparable and the evidence more usable for clinical decision-making. Our preliminary overview was based on the database of the Cochrane Collaboration, which provides information that can be used to investigate the overlap of review outcomes with trial outcomes but not vice versa.

The number of outcomes that trials report but are not included in systematic reviews, and the subsequent waste of research effort, is unknown and it was not addressed in our preliminary overview. If a sufficient lack of concordance were to exist between trial and review outcomes then the development of clinical guidelines could suffer. Consequently, primary researchers could lose motivation if significant amounts of their research were arguably wasted by failing to become incorporated into clinical guidelines. Funding bodies wishing to see a demonstrable impact for their money may likewise wish to ensure that trial outcomes are likely to be used by policy makers and guideline developers.

The primary aim of this study was to systematically investigate the overlap between outcomes reported in trials and those sought by Cochrane reviews and vice versa. Our secondary aim was to conduct a meta-epidemiological study into whether the pooled treatment effect differs between trials that are concordant with review primary outcomes versus those that are not.

RESULTS

The article selection process is displayed in figure 1. A summary of the Cochrane reviews, efficacy primary outcomes and meta-analyses is given in table 1 and the characteristics of the included trials are shown in supplement table 1. A total of 20/242 trials (8%) were excluded from the study due to the absence of full texts (7), not being published in a language native to the study's researchers (8) (English, German, Italian or Spanish) or containing no relevant information (5). In one review (Bamford et al., 2013) 11 of the trials evaluated were industry reports. All other trials were published in scientific journals. In the Cochrane reviews 23 outcomes were specified as primary (table 1, supplement table 2) and 37 as secondary (supplement table 3). Of note is that some texts contained information on more than one trial.

Our assessments of the trials' Risks of Bias ~~(risk of bias)~~ were given as "high" in 7 out of 10 reviews and "unclear" in the remaining three reviews. Risks of Bias ~~RoB~~ was noted to be particularly high across all trials for the blinding of participants and personnel (supplement table 4). It was also "high or unclear" in the allocation of sequence concealment, the blinding of trial outcome assessment and other potential threats to validity.

Lack of overlap between trial outcomes and review-sought outcomes

Figure 2 shows that of the 1086 outcomes identified amongst 220 trials, 742 (68%) were not included in the reviews. The median number of outcomes per trial was 4 (range 1 – 34). Of the 60 outcomes the reviews sought, 17 (28%) of these were not reported in any trial.

Missing outcomes

Supplement table 2 shows that 3/23 (13%) outcomes that reviews sought as primary were missing in every relevant trial. These were “participant-rated global improvement” (Chen et al. (2013) - Psoriasis), “improvement in sleep” (Ersser et al. (2014) - Atopic eczema) and “participant dis/satisfaction” (Kwok et al. (2012) - Warts). The mean presence in the trials of any specified review-sought primary outcome, irrespective of whether reported by trials as primary or secondary, was 45% (95% CI 32 to 59%).

14/37 (38%) secondary review outcomes were missing completely (supplement table 3). The mean presence in the trials of any specified review-sought secondary outcome, irrespective of whether reported by trials as primary or secondary, was 23% (95% CI 14 to 32%). The review with the most secondary trial outcomes (80%) was for pemphigus (Martin et al., 2009). The review with the least (21%) was for psoriasis (Chen et al., 2013).

Lack of concordance

Of the total 1086 trial outcomes that this study identified, those that the Cochrane authors defined as primary were seen 215 times. 31/215 (14%) of these were reported by the trials as primary. The outcomes that the reviews defined as secondary were present 129 times in the trials but only reported as secondary in 16 trials (12%).

Of the 23 outcomes sought as primary by the reviews 11 (48%) were not reported as such by the trials. 30/37 (81%) outcomes sought as secondary by the reviews were not reported as such by the trials.

The degree of concordance between reviews and trials agreeing on which outcomes should be primary or secondary is displayed in figure 3. The planned meta-epidemiologic analyses for the secondary objective of the study were not possible due to this lack of data.

The primary review outcome seen the most as a primary trial outcome ("Reduction in disease severity as measured [objectively] by a trained assessor") was present in 33% of trials assessed in the review by Ersser et al. (2014) (atopic eczema). The mean proportion of concordance between trial outcomes being sought as primary by reviews and being reported as primary by trials was 7% (95% CI 3 to 12%).

The secondary review outcome seen the most ("Response rates (partial and complete)") was reported as ‘secondary’ in 8% of trials assessed in the review by Sasse et al. (2007) (metastatic malignant melanoma). The mean concordance between secondary review outcomes and secondary trial outcomes was 2% (95% CI 0 to 4%).

Loss of potential data for systematic reviews

The Cochrane authors defined the outcomes they wished to consider in their reviews. In then selecting the trials to assess they implicitly established the maximum number of times they could have seen these outcomes. This potential figure can be derived as “number of outcomes sought by the review” multiplied by “number of trials included in the review”. Figure 4 illustrates how much of this data was indeed seen by the Cochrane reviews. Irrespective of whether trials reported outcomes as primary or secondary, the mean value of potential realized was only 35% (95% CI 27 to 43%). For trial outcomes specified as primary or secondary in concordance with the reviews' requirements the mean value of potential realized was 5% (95% CI 2 to 7%)).

The Cochrane reviewers frequently commented on the difficulties they faced due to absent or significantly heterogeneous trial outcomes and due to insufficient focus on patient-centered or quality-of-life matters (supplement table 5). In every single Cochrane review included in this study trial inconsistencies, lack of standardization and insufficient reporting were identified as major impediments for summarizing research evidence.

Outcome domains

An example of an outcome domain is "Quality of Life (QoL)" which would contain any outcome or measure which assessed QoL, irrespective of the actual instrument used. The trial outcomes considered by the reviews fell into several domains and these are shown in supplement table 6. This displays that despite a lack of concordance for many specific outcomes there was better agreement at the outcome domain level. For example, 27 trials into atopic eczema were assessed in Bamford et al. (2013); "Global disease severity" and "single symptom" were domains containing the primary outcomes defined by the Cochrane review and domains containing outcomes reported in 22 (85%) and 20 (77%) of trials respectively.

DISCUSSION

In evaluating the efficacy of interventions, evidence from randomized clinical trials represents scientifically rigorous evidence from a primary study design. Typically, clinical trials are designed to test the effect of exposure to a defined intervention versus control on specific primary efficacy or safety outcome(s). The trial design, statistical analysis employed and choice of sample size is focused on investigating the effect of an intervention on a primary outcome. The results of a trial can only be interpreted as confirmatory for this primary outcome (Hanson, 2008). Effects of the intervention on secondary trial outcomes have to be

interpreted as exploratory or hypothesis generating. The high level of evidence produced by clinical trials therefore applies predominantly to the primary outcome(s). From a methodological perspective, pooled evidence from confirmatory trials (i.e. trials with the same primary outcome as the review) may allow more valid conclusions regarding the efficacy of interventions than pooled evidence from exploratory trials (i.e. trials with a different primary outcome as the review). Meta-epidemiologic evidence on this critical research question is missing.

Main findings

Our study set out to analyze the degree of overlap between trial outcomes and those sought by reviews for a broad spectrum of dermatological conditions including psoriasis, melanoma, atopic eczema, pemphigus, seborrheic dermatitis, and cutaneous warts.

We identified a low degree of agreement between trial and review outcomes. This could indicate a vulnerability of systematic review conclusions. As the choice of outcomes in Cochrane reviews is designed to capture the key benefits and harms of treatments, this small overlap of review/trial outcomes could also suggest that many trials are not measuring the outcomes perceived to be the most important by patients, clinicians and other allied healthcare professionals.

Missing outcomes and loss of potential data for reviews

Our results demonstrate that a significant number of outcomes sought by reviews are missing in trials. This represents the first time such a substantial absence of data has been specifically identified in the dermatological field. The systematic review of any topic is impossible when

the required outcomes are not present in the published trials. Indeed due to this reason, and the heterogeneity of those trial outcomes that were actually reported, no meta-analysis was possible for 11/23 (48%) of the primary review outcomes. This represents a sizeable loss of possible conclusions to a review and to the potential subsequent evidence-based guidelines. 90% of reviews in our study were missing more than 50% of outcome data they would have had if every trial had reported every outcome the reviews wanted to assess. This finding is worse than that seen in a meta-analysis of general systematic reviews (Kirkham et al., 2010, Kirkham et al., 2013). In this, only 25% of reviews were missing more than 50% of potential data.

Limited usability of research effort

68% of all trial outcomes not being included in the reviews reflects how much research effort went no further than the original trial and leads to the question of how trialists decide which outcomes to use in their study.

Lack of concordance

The significant lack of concordance between reviews looking to assess specific outcomes as primary or secondary and trials investigating them as such had a substantial impact on the susbsequent reviews and meta-analyses. Review authors directly commented on this.

Mutual agreement of core outcomes, rather than being based solely on a review author's opinion, would particularly facilitate a relevancy informed by multiple viewpoints. This, in turn, may support better uptake and implementation by researchers of the concept.

Risk of Bias

Outcome reporting bias (ORB), the selection for publication of a subset of the originally recorded outcome variable (Williamson and Gamble, 2005), was more prevalent in our study than in a meta-analysis of general systematic reviews. This found 34% of assessed reviews contained at least one trial with high suspicion of ORB (Kirkham et al. (2010), recommended by Cochrane for systematic reviewers). The corresponding figure for our study was 70%.

Outcome domains

There was broader concordance between the trials and reviews with regards outcome domains than there was concordance for defined specific outcomes. This suggests that trialists and reviewers are, in general, considering similar areas of particular diseases rather than completely different aspects. As such, this implies that better concordance of specific outcomes is eminently achievable.

Limitations

We decided to analyse Cochrane reviews, because the Cochrane Handbook makes patient representation and involvement in review teams mandatory so that the choice of outcomes chosen in these reviews can be assumed to reflect the patient perspective. The 10 systematic reviews that we assessed represented only 15% of the 69 Cochrane Skin systematic reviews published during the time-period of sampling for our study. Since then 37 more have been published.

Our allocation of trial outcomes to systematic review outcome domains was performed iteratively. A uniform taxonomy was missing and allocation had the potential to be piecemeal.

Whether results apply to non-Cochrane reviews is unclear. For Cochrane, authors are encouraged to decide which primary review outcomes should be measured before looking at the data and to publish their intentions in a protocol. Non-Cochrane reviews may not do this and be more data driven, basing their choice on what is available. This could lead to better concordance but not necessarily better evidenced reviews. Evaluating the discordance in outcomes is particularly relevant for evidence-based practice.

Outcome identification in the primary trials was sometimes challenging, especially when the methods sections were not well developed. This difficulty may have led to undereporting of outcomes, if they were not clearly designated as such.

Implication for further research and perspective

Our study found a low degree of overlap between trial and review outcomes. This could facilitate weaknesses in systematic review conclusions due to insufficient data. Core Outcomes Sets (COS) may be a solution here and they are increasingly being developed to standardize outcomes across trials and to allow cross-trial comparisons. In dermatology, the harmonizing outcome measures in eczema (HOME) initiative pioneered COS development (Schmitt et al., 2015). Other groups have started COS development in other fields of dermatology, but none of these groups has yet implemented their COS.

We suggest that the main effort of future research in this field should be not only the development but also the implementation of COS in dermatology. The Cochrane Skin – Core Outcome Set Initiative (CS-COUSIN) has recently been established to improve and standardize outcome measurement in clinical trials and to make trial evidence more usable

(Kottner et al., 2018, Schmitt et al., 2016). Further aims are to develop clinically relevant and patient-centered sets of dermatological trial outcomes and improve the quality and interpretability of systematic reviews. Ultimately, the objective is to facilitate dermatological research delivering a tangible clinical impact for patients through evidence-based healthcare. The implementation of COS requires high acceptability of all stakeholder groups including key clinicians, researchers, patient representatives, but also regulators (FDA, EMA), pharmaceutical industry, and journal editors. Journal editors have been for example involved in the first HOME Delphi study published in the JID (Schmitt et al., 2011). Ideally, journal editors should encourage the application of existing COS in their guidelines for authors. Funders of clinical trials already do so in the UK and in Germany.

The CS-COUSIN initiative will primarily achieve this by supporting research groups as they develop a COS. Specifically, CS-COUSIN provides expert and ongoing peer-based methodological advice (e.g., regarding outcome measurement (Grinich et al., 2018)), a development pathway and administrative support for the lifespan of the COS projects. Its structured and experienced approach can also help reduce bias risk. Core outcomes and, crucially, measurement instruments are currently in progress for 16 disease areas (including acne, chronic spontaneous urticaria, chronic wounds, atopic & hand eczema, melanoma, rosacea, hidradenitis suppurativa, nail psoriasis and basal cell carcinomas). CS-COUSIN can further help to reduce the risk of bias which in our study was also worse than in others. Specifically, the impact of selective reporting of outcomes seen in Outcome Reporting Bias could be reduced (Williamson et al., 2012). This has been shown to hamper systematic reviews significantly (Kirkham et al., 2010).

The CS-COUSIN initiative is constantly open to new members looking to develop a COS, join a current project group, use a specific COS, receive methodological advice or join as a patient-representative. To discover more, and how we could help you make COS in dermatology a reality, please scan the website and twitter QR codes in figure 5 or visit our website (www.cs-cousin.org).

MATERIALS AND METHODS

This study was an investigation of a random sample of systematic Cochrane Skin Group Reviews and the component trials.

Protocol and registration

A review protocol was developed and published at PROSPERO in September 2015 (CRD42015025005).

Data source and data collection

A random sample of 10 systematic reviews of the Cochrane Skin Group published until January 2015 and their component trials. A random sample was drawn, using the statistical software "R", from all 69 CSG reviews in dermatological diseases published until January 2015. The 10 reviews included 242 primary studies (range: 7 to 85 studies). Only trials written in English, German, Italian or Spanish were included.

Data extraction

Where possible, Cochrane provided clinical trial data from the ARCHIE database (https://archie.cochrane.org/). This consisted of data originally extracted by the systematic

review team from the trials they included during their reviewing process. Independent reviewers extracted data at both trial and systematic review level and data were double extracted.

Review global-data extraction

The following information was extracted from each review; whether a meta-analysis was performed, the reasons why no meta-analysis was performed as applicable, specifically "heterogeneity" in outcome measurement across trials as the reason, the number of trials included in the meta-analysis of the primary efficacy outcome of the systematic review and the number of trials not included in the meta-analysis of the primary efficacy outcome (total and due to no assessment of primary review outcome in underlying trials).

Trial global-data extraction

For each trial we extracted published data on general characteristics (year, geographical bias (assessed by the Cochrane risk of bias instrument), outcomes (i.e. primary, secondary, not specified), and results (with respect to the primary outcome the number of patients randomized and the number of events per group).

If outcomes were not explicitly classified as primary or secondary, we assumed an outcome was a primary outcome when sample size calculations were based on this outcome. Otherwise, the outcome was classified as "not specified". If there was a reference to a study protocol in the trial, we also obtained this to extract the relevant data.

Outcome extraction

For each trial and review outcome we extracted the outcome domain(s), measurement instrument or outcome definition, number of participants and events in both the intervention and the control group (for reviews only applicable if meta-analysis has been conducted) and whether the outcome was the primary or secondary outcome or if there was no information in this respect.

Protocol deviation

We had planned to compare pooled outcomes from meta-analysis of all trials to the pooled effects of (1) component trials with the primary review outcome used as primary, (2) secondary, or (3) unspecified outcome with the 2-step meta-epidemiologic approach (Savovic et al., 2012). The planned meta-epidemiologic analyses, however, for the secondary objective of the study became unfeasible due to a lack of data.

Data collection process

Outcomes were extracted in duplicate independently from the original publications by two reviewers using a standardized electronic template in MS Access. Then both data sets were compared and all areas of disagreements were resolved.

DECLARATION OF INTEREST

CA received consultancy fees from Dr Wolff GmbH for consulting on patient-reported outcomes in skin disease.

ACKNOWLEDGMENTS

We thank the Cochrane Skin management team for providing access to ARCHIE data. We also thank Jacqui Moreau for input into the study protocol.

REFERENCES

Reference List

Bamford JT, Ray S, Musekiwa A, van Gool C, Humphreys R, Ernst E. Oral evening primrose oil and borage oil for eczema. Cochrane Database Syst Rev 2013(4):CD004416.

Bath-Hextall FJ, Jenkinson C, Humphreys R, Williams HC. Dietary supplements for established atopic eczema. Cochrane Database Syst Rev 2012(2):CD005205.

Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. The Journal of rheumatology 1998;25(2):198-9.

Boers M, Kirwan, J.R.,Tugwell, P. The OMERACT Handbook. OMERACT; 2014.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet (London, England) 2009;374(9683):86-9.

Chen X, Yang M, Cheng Y, Liu GJ, Zhang M. Narrow-band ultraviolet B phototherapy versus broad-band ultraviolet B or psoralen-ultraviolet A photochemotherapy for psoriasis. Cochrane Database Syst Rev 2013(10):CD009481.

Chi CC, Kirtschig G, Baldo M, Brackenbury F, Lewis F, Wojnarowska F. Topical interventions for genital lichen sclerosus. Cochrane Database Syst Rev 2011(12):CD008240.

Eekhof JA, Van Wijk B, Knuistingh Neven A, van der Wouden JC. Interventions for ingrowing toenails. Cochrane Database Syst Rev 2012(4):CD001541.

Ersser SJ, Cowdell F, Latter S, Gardiner E, Flohr C, Thompson AR, et al. Psychological and educational interventions for atopic eczema in children. Cochrane Database Syst Rev 2014(1):CD004054.

Grinich EE, Schmitt J, Kuster D, Spuls PI, Williams HC, Chalmers JR, et al. Standardized reporting of the Eczema Area and Severity Index (EASI) and the Patient-Oriented Eczema Measure (POEM): a recommendation by the Harmonising Outcome Measures for Eczema (HOME) Initiative. The British journal of dermatology 2018;179(2):540-1.

Hanson B. The art of choosing sound study endpoints. Injury 2008;39(6):656-8.

Kastarinen H, Oksanen T, Okokon EO, Kiviniemi VV, Airola K, Jyrkka J, et al. Topical anti-inflammatory agents for seborrhoeic dermatitis of the face or scalp. Cochrane Database Syst Rev 2014(5):CD009446.

Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ (Clinical research ed) 2010;340:c365.

Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews?--a survey of the Co-ordinating Editors of Cochrane Review Groups. Trials 2013;14:21.

Kottner J, Jacobi L, Hahnel E, Alam M, Balzer K, Beeckman D, et al. Core outcome sets in dermatology: report from the second meeting of the International Cochrane Skin Group Core Outcome Set Initiative. The British journal of dermatology 2018;178(4):e279-e85.

Kwok CS, Gibbs S, Bennett C, Holland R, Abbott R. Topical treatments for cutaneous warts. Cochrane Database Syst Rev 2012(9):CD001781.

Martin LK, Werth V, Villanueva E, Segall J, Murrell DF. Interventions for pemphigus vulgaris and pemphigus foliaceus. Cochrane Database Syst Rev 2009(1):CD006263.

Sasse AD, Sasse EC, Clark LG, Ulloa L, Clark OA. Chemoimmunotherapy versus chemotherapy for metastatic malignant melanoma. Cochrane Database Syst Rev 2007(1):CD005413.

Savovic J, Jones H, Altman D, Harris R, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health technology assessment (Winchester, England) 2012;16(35):1-82.

Schmitt J, Apfelbacher C, Spuls PI, Thomas KS, Simpson EL, Furue M, et al. The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. The Journal of investigative dermatology 2015;135(1):24-30.

Schmitt J, Deckert S, Alam M, Apfelbacher C, Barbaric J, Bauer A, et al. Report from the kick-off meeting of the Cochrane Skin Group Core Outcome Set Initiative (CSG-COUSIN). The British journal of dermatology 2016;174(2):287-95.

Schmitt J, Langan S, Stamm T, Williams HC, Harmonizing Outcome Measurements in Eczema Delphi p. Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. J Invest Dermatol 2011;131(3):623-30.

Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. PLoS medicine 2008;5(4):e96.

Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.

Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. Statistics in medicine 2005;24(10):1547-61.

Tables

Table 1: Summary of included Cochrane reviews, primary efficacy outcomes, and realization meta-analyses

Cochrane Review **(Number of trials included in this study / number in review)**	**Mention of primary outcome in abstract or plain language summary**	**Primary Review Outcome(s)**	**If sub-group or meta-analysis was conducted n of trials included / not included**	**Reason for non-inclusion of trials in meta-analysis**	**If no meta-analysis was conducted** **Reason**
Bamford et al. (2013) (atopic eczema) (26/27)	In plain language summary	Primary efficacy outcome I: Global degree of improvement in symptoms and signs as rated by participant or medical doctor.	**Participant**: 12/27 reported primary efficacy outcome, 7/12 included in meta- analysis. 2 subgroup analyses (1 included 2 trials, 1 included 5 trials). **Clinician**: 19/27 reported primary efficacy outcome 8/19 included in meta-analysis. 2 subgroup analyses (1 included 3 trials, 1 included 5 trials)		Trials reported this outcome, but in different ways.
		Primary efficacy outcome II: Improvement in quality of life.	No meta-analysis		Only 2 trials reported the primary outcome but these could not be pooled as they compared different interventions
Bath-Hextall et al. (2012) (atopic eczema) (10/11)	In plain language summary	Primary efficacy outcome I: Degree of long-term (over six months) control, such as reduction in number of flares or reduced need for other treatments.	No meta-analysis		For 9/10 trials no data was available for the primary outcomes
		Primary efficacy outcome II: Short-term (within six weeks). Changes in participant-rated or parent-rated symptoms of atopic eczema, such as pruritus (itching) or sleep loss.			
Chen et al. (2013) (psoriasis) (13/13)	In abstract	Primary efficacy outcome I: Participant-rated global improvement.	No meta-analysis		None of the 13 included trials addressed this primary outcome.
		Primary efficacy outcome II: Percentage of participants reaching Psoriasis Area and Severity Index (PASI) 75.	No meta-analysis		Only 2/13 trials with different interventions addressed this primary outcome.
		Primary efficacy outcome III: Clearance rate, defined as no lesions of psoriasis or minimal residual activity (MRA)	10 from 13 studies assessed this primary outcome; 4 studies included in meta-analysis: 2 meta-analysis with each 2 studies each subgroup 1: 2/1 subgroup 2: 2/0	none or only one study per intervention, once because of heterogeneity between the studies	
Chi et al. (2011) (lichen sclerosus (genital) (7/7)	No	Primary efficacy outcome I: Participant-rated improvement or remission of symptoms (in terms of quality of life, pain, itching, and less pain with intercourse).	6 of 7 studies assessed the outcome 2 studies included in 1 meta-analysis	1 study did not evaluate this outcome.	
		Primary efficacy outcome II: Investigator-rated global degree of improvement (in terms	7 of 7 studies assessed the outcome no meta-analysis		

		of pallor, purpura, hyperkeratosis, ulceration, erosion, erythema,sclerosis, and scarring).	NB: Paper reports in text that "All seven included studies reported investigator-rated improvement of gross appearance" however, again, variable language used in meta analysis for this outcome.		
Eekhof et al. (2012) (ingrowing toenails) (21/24)	In abstract and plain language summary	Primary efficacy outcome I: Relief of symptoms	no meta-analysis		
		Primary efficacy outcome II: Recurrence	16 from 24 studies assessed primary outcome recurrence; 1 meta-analyses subgroup 1: 2/1		
		Primary efficacy outcome III: Regrowth (including nail spicules/nail spikes)	no meta-analysis		
Ersser et al. (2014) (atopic eczema) (9/10)	In abstract	Primary efficacy outcome I: participant-rated global assessment	No meta-analysis		None of the 9 included studies addressed this primary outcome.
		Primary efficacy outcome II: reduction in disease severity, determined by an assessor using an objective measure	No meta-analysis 8 / 9 studies assessed severity (6x SCORAD)		1. methodological weaknesses in the selected studies; 2. heterogeneity of the outcome measures; and 3. the heterogeneous nature of the interventions.
		Primary efficacy outcome III: improvement in sleep	No meta-analysis		None of the 9 included studies addressed this primary outcome
		Primary efficacy outcome IV: improvement in quality of life (or reduction in distress of the child and parent)	No meta-analysis 7 from 9 studies assessed QoL		1. methodological weaknesses in the selected studies; 2. heterogeneity of the outcome measures; and 3. the heterogeneous nature of the interventions.
Kastarinen et al. (2014) (seborrheic dermatitis (face and scalp) (35/36)	In abstract	Primary efficacy outcome I: Total clearance (total resolution of symptoms), evaluated by an outcome assessor [expressed as a % of people treated].	27 /36 trials assessed the outcome total clearance 15 trials included in 4 meta analyses	1 study excluded as mode of application was different from all other trials. 1 study excluded as clearance defined as ≥75% which was less than in other included trials.	
		Primary efficacy outcome II: Disease severity scores for scaling, pruritus, or erythema at the end of treatment as (evaluated by participant self-report, outcome assessor, or both).	28/ 36 trials assessed one of the outcome (scaling, pruritus or erythema) 7 trials included in 3 meta analyses		
Kwok et al. (2012) (common warts,	No [Efficacy mentioned, but no	Primary efficacy outcome I: Clinical cure (defined as complete disappearance of elevated/warty skin) at end of treatment period.	37/85 assessed this primary outcome. Meta-analysis:	The heterogeneity of the trials made it difficult to	

cutaneous warts (non-genital)) (70/85)	further details given]		35 included in meta-analysis: 10 meta-analysis plus subgroup-analysis with meta-analysis	perform statistical pooling of the data.
		Primary efficacy outcome II + III: Participant satisfaction/dissatisfaction / Quality of Life.	No meta-analysis	
Martin et al. (2009) (pemphigus vulgaris and pemphigus foliaceus) (10/11)	No	Primary efficacy outcome I: The proportion of participants achieving remission (defined as the absence of lesions or the presence of transient new lesions that heal within one week, while the personis receiving minimal therapy).	3 /11 trials evaluated the specified outcome measure 2 trials included in meta analysis (but with the data from only one study, because the other is not estimable)	
Sasse et al. (2007) (metastatic malignant melanoma) (18/18)	No	Primary efficacy outcome I: Overall survival [number of participants alive at end of trial].	8/18 trials evaluated the outcome overall survival, all 8 trials were included in a meta-analysis	