Is there an association between study size and reporting of study quality in dermatological clinical trials? A meta-epidemiological review.

S Ratib[1] PhD, SR Wilkes[1] PhD , MJ Grainge[2] PhD, KS Thomas[1] PhD, C Tobinska[3], HC Williams[1] PhD

Affiliations
[1]Centre of Evidence Based Dermatology, University of Nottingham, UK

[2]Division of Epidemiology and Public Health, University of Nottingham, UK

[3]University of Nottingham, Medical School, Nottingham, UK

Dear Editor,

Clinicians engaged in evidence-based medicine often have to base treatment decisions on randomised clinical trials (RCTs) with small sizes.[1] The choice of sample size is influenced by a number of factors including power, magnitude of difference anticipated, alpha, study design and whether the objective of the trial is equivalence or superiority. Based on our experience of reviewing hundreds of dermatological clinical trials, we hypothesized that sample size could be a crude surrogate of study quality in that larger studies are of generally better quality than smaller ones. Given the substantial evidence that some specific trial characteristics such as allocation concealment may influence treatment effect,[1,2,3] knowledge of study quality before implementing trial results into clinical practice is important. We therefore aimed to examine if there is an association between sample size and risk-of-bias assessment in RCTs included in published Cochrane skin reviews. The Cochrane risk-of-bias tool focusses on whether key items, known to be empirically associated with distortion of treatment effect, are reported or not.[3,4] We therefore refer to 'reporting of study quality' as a proxy for formal study quality.

To perform the assessment, we identified all Cochrane Skin Group reviews and updates published between 2010 and 2014. For each of the included trials we extracted: sample size (total number randomised), study design (parallel/non-parallel), type of study (full/pilot/feasibility) and year of publication. Using the risk-of-bias tables included in the individual Cochrane reviews, two authors independently extracted the Cochrane risk-of-bias assessment (low, high or unclear) for the following six quality criteria: (i) methods for sequence generation, (ii) allocation concealment, (iii) blinding of participants and personnel, (iv) blinding of outcome assessment, and (v) incomplete outcome data.

We used the Kruskal-Wallis test to determine if there was a statistically significant association between sample size and risk-of-bias category, for each quality criterion. The Mann-Whitney-U test was performed to assess each of the three pairwise comparisons. Given within-patient and cross-over trials require fewer patients than parallel design trials; we stratified the results by study design and repeated the analysis. We also performed a *post-hoc* analysis by stratifying the data by decade of trial publication (1966-1975, 1976-1985, 1986-1995, 1996-2005, 2006-2014) to assess whether the nature of the association changed over time. Finally, we determined if the proportion of trials with unclear risk-of-bias for each quality criterion changed over time using the chi-squared test. We adjusted for multiple testing using a Bonferroni correction.

In total, we identified 24 new reviews and 10 updates and 1,127 included trials within these. The median sample size was 72 [IQR 40-151]. The majority of trials were parallel design (n=928, 82.3%). Of the 1,127, 4 were feasibility and 18 were pilot trials.

For random sequence generation and allocation concealment the majority of trials were assigned unclear risk-of-bias (64.3% and 79% respectively)(Table 1). In contrast, for blinding

participants/personnel, blinding-outcome assessment and incomplete outcome data approximately half of the trials had a low risk-of-bias (43.5%, 44.7% and 56% respectively).

Overall, average sample size did not differ significantly between the three risk-of-bias categories for any of the quality criteria (Table 1). For all quality criteria apart from incomplete outcome data, trials with low risk-of-bias had a larger sample size than those at high risk-of-bias. Comparing median sample size of low vs. unclear trials, there were also no statistically significant differences for any of the quality criteria. Similar findings applied to both parallel and non-parallel studies.

During the first (1966-75) and last (2006-14) decades, there was no statistically significant association between sample size and risk-of-bias category for any of the quality criteria. In the intermittent years there were some significant associations, however after accounting for multiple testing these were no longer statistically significant.

The proportion of trials with unclear risk-of-bias reporting decreased substantially in the last two decades for (i) methods for sequence generation (62.6%vs.46.2%%;p<0.001), (ii) allocation concealment (79.3%vs.69.2%;p<0.003), and (iii) blinding of participants/personnel (29.3%vs.21.6%;p=0.021).

Our study failed to show any statistically significant association between study size and quality, although the direction was as we expected for most quality criteria. The high degree of unclear risk-of-bias for some items made it very difficult to come to any firm conclusions. Our study underlines the critical importance of improving the quality of reporting of dermatological trials by applying reporting guidelines such as using CONSORT 2010. Large and statistically significant associations between sample size and risk-of-bias could still exist if all key items had been reported properly.[1,5] It is reassuring that the proportion of criteria with unclear risk-of-bias has declined over the last decade, yet the overall quality of reporting is still poor. It is

possible that other key (but as yet unknown) factors are critical for determining study quality as opposed to reflecting quality of reporting. We conclude sample size cannot be used as a short-cut for determining reporting of study quality as assessed by current methods to assess risk-of-bias.

Table 1 Median sample size by risk of bias for each quality criterion

Quality criteria	Risk of bias			Kruskal Wallis p-value
	Low	High	Unclear	
Allocation concealment n (%)	**201 (17.8%)**	**36 (3.2%)**	**890 (79%)**	
median sample size [IQR]	84 [40, 192]	63.5 [48.5, 100]	70 [39, 149]	0.13
Random sequence generation n(%)	**390 (34.6%)**	**11 (1%)**	**726 (64.3%)**	
median sample size [IQR]	72 [36, 165]	50 [37, 140]	72 [40, 149]	0.79
Blinding-participants/personnel n(%)	**490 (43.5%)**	**290 (25.7%)**	**347 (30.8%)**	
median sample size [IQR]	71 [34, 171]	69 [45, 126]	75 [40, 147]	0.94
Blinding-outcome assessment n(%)	**504 (44.7%)**	**213 (18.9%)**	**410 (36.4%)**	
median sample size [IQR]	71 [34, 176]	66.5 [40.5, 122]	75 [40, 142]	0.82
Incomplete outcome data n(%)	**630 (56%)**	**267 (23.7%)**	**229 (20.3%)**	
median sample size [IQR]	70 [36, 150]	80 [47, 155]	70 [40, 150]	0.23

For each quality criterion, the number and proportion of trials in each risk-of-bias category, and the median sample size [inter-quartile range; IQR] per risk-of-bias category. The Kruskal-Wallis test determines any statistically significant association between sample size and risk of bias category.

References

1. Gludd LL. Bias in clinical intervention research. *Am J Epidemiol. 2006;* 163: 493-501.
2. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. Jama. 1999 Sep 15;282(11):1054-60. PubMed PMID: 10493204.
3. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273:408-412
4. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. Bmj. 2009;339:b4012. PubMed PMID: 19841007. Pubmed Central PMCID: 2764034.
5. Williams HC. Cars, CONSORT 2010, and clinical practice. Trials. 2010;11:33. PubMed PMID: 20334635. Pubmed Central PMCID: 2857833.