PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	RANDOMISED CONTROLLED TRIAL OF TOPICAL KANUKA HONEY FOR THE TREATMENT OF ACNE
AUTHORS	Semprini, Alex; Braithwaite, Irene; Corin, Andrew; Sheahan, Davitt; Tofield, Chris; Helm, Colin; Montgomery, Barney; Fingleton, James; Weatherall, Mark; Beasley, Richard

VERSION 1 - REVIEW

REVIEWER	Steven Feldman Wake Forest University School of Medicine Winston-Salem, NC USA I have had support from many companies, including some that make other products for acne.
REVIEW RETURNED	25-Jul-2015

GENERAL COMMENTS	This is a well written, well organized, randomized controlled trial that did not show any strong benefit. Publishing such trials is valuable. 1. Wordiness can be reduced. Whole sentences are not needed to let the reader know about tables. For example, considering changing "The flow of participants is shown in Figure 1. There were 68 participants randomised to each treatment arm. There were similar proportions of males and females the duration of acnewas similar in the two arms. Table 2 describes the participants," to "There were 68 participants randomised to each treatment arm (figure 1). There were similar proportions of males and females the duration of acne was similar in the two arms (Table 2)." 2. The Results text should be much shorter, presented primarily in tables. 3. I did not understand what the "exponents" were in the Results. 4. IGA can be a very insensitive measure of improvement. Lesion counts may be more sensitive, but they, too, did not achieve statistical significance. 5. I doubt that adherence "is likely dependent on the perceived acne severity." Adherence in teenagers with acne is poor, even when they are bothered by their acne. 6. A future study might benefit from inclusion of a positive control, to show that the methods of the study would be capable of detecting a benefit if there was one.

REVIEWER	Hywel Williams University of Nottingham UK
REVIEW RETURNED	29-Oct-2015

GENERAL COMMENTS	Overall, I enjoyed reading this submission. It was clearly written, well

	laid out, and was accompanied by a protocol, prospective clinical trial registration, and a CONSORT checklist – how refreshing. It addressed a much needed topic – that of addressing uncertainties for potentially useful topical treatments for acne that may be popular with patients. Most of the studies that I have read on complementary or alternative treatments for acne tend to be small and very poorly reported, so this study is welcome. What I also liked about the study is that despite not finding any evidence of a convincing treatment benefit, the study is very openly written without spin or framing bias, despite the potential influence of the study funders. I also like some aspects of the discussion such as the problem of floor effects when investigating participants with acne in primary care where most patients with acne present. I also like the brave decision to choose a primary outcome measure that reflects a decent magnitude of benefit. So many of the acne studies look at 20% reduction in inflammatory acne lesions as a primary outcome, but I have yet to meet a teenager who comes into my clinic who is pleased with such a reduction (even though it might be statistically significant in a large study). The authors have chosen the sort of magnitude of improvement that would influence patients and clinicians, so well done, even though you have not found a clear treatment benefit. My concerns below are mainly minor, and can easily be addressed by minor revisions. 1. I am always suspicious of RCTs that end up dividing the study population into exactly the same number for both treatment arms (68). The probability of this occurring with true simple randomisation is very small as previous studies by Doug Altman have shown. Perhaps the authors used blocking or some other method to ensure equal sample size, or was it purely chance? 2. I didn't entirely agree that it was not possible to blind participants. You could have used partial blinding with some other inert substance in order to decrease performance bias. As it happened, we cannot really suggest that performance bias had much to do with the results given the lack of efficacy, but you could bear this in mind for future studies. 3. No mention is made of co-treatment. Did participants use other treatments that were not forbidden in your protocol, and if so, how much? I realise that such data is difficult to collect, but it is a potential hazard for an unblinded intervention. 4. The decision to use odds ratios as your summary estimate did not sit well with me given that you were expecting common event rates. You based your sample size calculation on an absolute risk reduction, and I think the abstract is much clearer if you simply state the difference in primary outcome percentages between the two groups along with corresponding 95% confidence intervals. 5. No mention is made in the analysis section on whether you planned to do an intention-to-treat analysis or per-protocol analysis. You have clearly done a per-protocol analysis which is fine for an efficacy study, but you need to state it in the methods analysis section. 6. Please explain how you asked your second investigators to

	measure the change in IgA at 12 weeks compared with baseline. I very much doubt if the investigators could accurately recall what the patients looked like previously, but it is not clear whether they were presented with their baseline scores (which could induce some degree of bias as they knew which patients were before and after), or whether they simply recorded the IgA as a point estimation (which sounds better) with a point estimation at 12 weeks subtracted from the baseline. Please clarify exactly what was done here. 7. The discussion is a bit of a jumble and would really benefit from some structured sub-headings, eg. Main findings, Strengths and study limitations, Implications for clinical practice, and Implications for further research. 8. A little bit of 'trying to save face' crept into the discussion by pointing to insignificant trends. You mention that the subject rated improvement to a VAS was better for the honey treatment, whereas the result for the between-treatment differences for the active versus control treatment were not significant. I would tone this bit down as it distracts from the otherwise open and honest presentation of the results.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Steven Feldman
Institution and Country: Wake Forest University School of Medicine, Winston-Salem, NC USA

This is a well written, well organized, randomized controlled trial that did not show any strong benefit. Publishing such trials is valuable.

1. Wordiness can be reduced. Whole sentences are not needed to let the reader know about tables. For example, considering changing "The flow of participants is shown in Figure 1. There were 68 participants randomised to each treatment arm. There were similar proportions of males and females the duration of acne was similar in the two arms. Table 2 describes the participants," to "There were 68 participants randomised to each treatment arm (figure 1). There were similar proportions of males and females the duration of acne was similar in the two arms (Table 2)."

Answer: Thank you for your suggestions. Wordiness has been reduced with particular focus surrounding the description of tables as suggested.

2. The Results text should be much shorter, presented primarily in tables.

Answer: The multiplicity of secondary outcome variables made the inclusion tabulation of the statistics within tables non- viable in terms of report real estate. This was compounded by the need for explanation of each and we therefore made the decision to briefly dedicate a sentence to each.

3. I did not understand what the "exponents" were in the Results.

Answer: Thank you for this comment request for clarification. We have stated in the methods section that the exponent of the difference in logarithms can be interpreted as the ratio of mean values. We have further expanded this in the results section by replacing 'exponent' with 'ratio of mean lesion counts by exponentiation'.

4. IGA can be a very insensitive measure of improvement. Lesion counts may be more sensitive, but

they, too, did not achieve statistical significance.

Answer: Thank you for your comment, we have made no changes to the manuscript with respect to this point and concur with this interpretation.

5. I doubt that adherence "is likely dependent on the perceived acne severity." Adherence in teenagers with acne is poor, even when they are bothered by their acne.

Answer: Thank you, we have amended this point in the following text to reflect this comment: "

'We considered that the motivation of an individual to adhere to such a treatment and record keeping regimen, and subsequent quality of data collected, is likely dependent on the perceived acne severity however, it is also possible that adherence in teenagers with acne is poor, even when they are bothered by their acne."'

6. A future study might benefit from inclusion of a positive control, to show that the methods of the study would be capable of detecting a benefit if there was one.

Answer: Thank you for this suggestion.

Reviewer: 2
Reviewer Name: Hywel Williams
Institution and Country: University of Nottingham, UK.

Overall, I enjoyed reading this submission. It was clearly written, well laid out, and was accompanied by a protocol, prospective clinical trial registration, and a CONSORT checklist – how refreshing. It addressed a much needed topic – that of addressing uncertainties for potentially useful topical treatments for acne that may be popular with patients. Most of the studies that I have read on complementary or alternative treatments for acne tend to be small and very poorly reported, so this study is welcome.

What I also liked about the study is that despite not finding any evidence of a convincing treatment benefit, the study is very openly written without spin or framing bias, despite the potential influence of the study funders. I also like some aspects of the discussion such as the problem of floor effects when investigating participants with acne in primary care where most patients with acne present. I also like the brave decision to choose a primary outcome measure that reflects a decent magnitude of benefit. So many of the acne studies look at 20% reduction in inflammatory acne lesions as a primary outcome, but I have yet to meet a teenager who comes into my clinic who is pleased with such a reduction (even though it might be statistically significant in a large study). The authors have chosen the sort of magnitude of improvement that would influence patients and clinicians, so well done, even though you have not found a clear treatment benefit.

Answer: Thank you for these comments; we note in relation to the request from the editor about the sample size calculation that this reviewer considers the effect size was appropriate.

My concerns below are mainly minor, and can easily be addressed by minor revisions.

1. I am always suspicious of RCTs that end up dividing the study population into exactly the same number for both treatment arms (68). The probability of this occurring with true simple randomisation is very small as previous studies by Doug Altman have shown. Perhaps the authors used blocking or

some other method to ensure equal sample size, or was it purely chance?

Answer: Thank you, we in fact blocked by the whole 136 so that by design there were the same number of participants in each treatment arm.

2. I didn't entirely agree that it was not possible to blind participants. You could have used partial blinding with some other inert substance in order to decrease performance bias. As it happened, we cannot really suggest that performance bias had much to do with the results given the lack of efficacy, but you could bear this in mind for future studies.

Answer: Thank you for this point. Unfortunately the lack of participation blinding is unavoidable due to the Honevo both smelling and tasting of honey. Producing an inert substance to add to the control would not be possible as a high concentration of honey would be required to match these characteristics, essential as the participants are aware of the active treatment from the consenting process.

3. No mention is made of co-treatment. Did participants use other treatments that were not forbidden in your protocol, and if so, how much? I realise that such data is difficult to collect, but it is a potential hazard for an unblinded intervention.

Answer: While participants were asked about new medication use via a standard concomitant medication form at each follow up visit we recognize that this would not necessarily include all topical preparations and/or undeclared use of any such medication. For the purposes of analysis, only the forbidden medications deemed to influence the participants underlying acne, as listed within the protocol were recorded.

4. The decision to use odds ratios as your summary estimate did not sit well with me given that you were expecting common event rates. You based your sample size calculation on an absolute risk reduction, and I think the abstract is much clearer if you simply state the difference in primary outcome percentages between the two groups along with corresponding 95% confidence intervals.

Answer: Thank you for this comment. The decision to use logistic regression and estimate an odds ratio involved was the subject of much discussion amongst the research team. This was because although we felt the IGA was best treated as an ordinal scale variable and analysed using a proportional odds model (which produces an estimate of the common odds ratio between any ordered division of best versus worst on the instrument) we thought that clinicians would find the single division of best versus worst (at a boundary point of 2) more clinically meaningful. For consistency though, with the secondary analysis by proportional odds regression, we had then decided to preserve the single odds ratio estimate. As can be seen there is some consistency (in the same direction) between the proportional odds estimates of association with a better outcome of 1.4 and 2 with the 'head-line' estimate of 4.2 with the single dichotomous cut-point; but that we have paid a price with the lack of precision for the single cut-point estimate (very wide confidence intervals because of loss of information). Our preference is to preserve the estimates as written, but that because we have provided detailed summary information that may interested readers we could easily estimate the difference in proportions themselves, but would be happy to take the editors' advice about the best way forward here.

5. No mention is made in the analysis section on whether you planned to do an intention-to-treat analysis or per-protocol analysis. You have clearly done a per-protocol analysis which is fine for an efficacy study, but you need to state it in the methods analysis section.

Answer: Thank you for alerting us to this, the per protocol analysis has now been specified.

6. Please explain how you asked your second investigators to measure the change in IgA at 12 weeks compared with baseline. I very much doubt if the investigators could accurately recall what the patients looked like previously, but it is not clear whether they were presented with their baseline scores (which could induce some degree of bias as they knew which patients were before and after), or whether they simply recorded the IgA as a point estimation (which sounds better) with a point estimation at 12 weeks subtracted from the baseline. Please clarify exactly what was done here.

Answer: Thank you for this important comment. We have clarified the text to reflect that the change in IgA at 12 weeks compared with baseline was calculated at the analysis stage, from point estimates obtained at each visit.

7. The discussion is a bit of a jumble and would really benefit from some structured sub-headings, eg. Main findings, Strengths and study limitations, Implications for clinical practice, and Implications for further research.

Answer: Thank you. The discussion has been amended and expanded accordingly in line with both yours and the editor's suggestion.

8. A little bit of 'trying to save face' crept into the discussion by pointing to insignificant trends. You mention that the subject rated improvement to a VAS was better for the honey treatment, whereas the result for the between-treatment differences for the active versus control treatment were not significant. I would tone this bit down as it distracts from the otherwise open and honest presentation of the results.

Answer: Thank you for this comment. The subject rated VAS improvement was statistically significant at both week 4 (9.6 (3.7 to 15.6), P = 0.002) and week 12 (11.0 (4.4 to 17.6) P = 0.001) favouring honevo and this is stated within a paragraph reinforcing lack of signal within the secondary outcomes in general, along with further caution in interpretation in terms of potential type I error and assessment bias.

We have also ensured that our reference to a trend for improvement when IGA was treated as an ordinal variable is further qualified as being statistically insignificant within the discussion.

VERSION 2 – REVIEW

REVIEWER	Steven Feldman Wake Forest University School of Medicine, USA I have had research, speaking and consulting support from many companies with products for skin disease, but I don't think any are relevant to this paper.
REVIEW RETURNED	09-Dec-2015

GENERAL COMMENTS	The authors addressed the issues that were raised previously. There is extensive use of brand names in the paper. Typically, I think brand names usually should appear once and only once-- in the Methods section.

REVIEWER	Hywel Williams Centre of Evidence-based dermatology at the university of nottingham
REVIEW RETURNED	09-Dec-2015

GENERAL COMMENTS	Well done for revising so nicely. The structure and balance and clarity of the study are now all improved.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Steven Feldman
Institution and Country: Wake Forest University School of Medicine, USA

Please leave your comments for the authors below

The authors addressed the issues that were raised previously.

There is extensive use of brand names in the paper. Typically, I think brand names usually should appear once and only once-- in the Methods section.

Thank you, the brand names have been replaced with ‘honey product’ and ‘control’ with initial product definitions in the abstract and methods sections.

Reviewer: 2
Reviewer Name: Hywel Williams
Institution and Country: Centre of Evidence-based Dermatology at the University of Nottingham, UK

Please leave your comments for the authors below

Well done for revising so nicely. The structure and balance and clarity of the study are now all improved.