Author's response to reviews

Title: Evaluation of a bespoke training to increase uptake by midwifery teams of NICE Guidance for membrane sweeping to reduce induction of labour: a stepped wedge cluster randomised design

Authors:

Sara Kenyon (s.kenyon@bham.ac.uk)
Sophie Dann (sophiedann@gmail.com)
Lucy Hope (l.hope@worc.ac.uk)
Paula Clarke (PAULA.CLARKE@bwnft.nhs.uk)
Amanda Hogan (amandahogan1@nhs.net)
David Jenkinson (D.J.Jenkinson@warwick.ac.uk)
Karla Hemming (k.hemming@bham.ac.uk)

Version: 1 Date: 07 Feb 2017

Author’s response to reviews:

We are most grateful for the opportunity to respond to reviewers comments and have done so in the table below. We hope that you will review these and the revised paper and further consider the paper for publication.

Reviewer #1: Abstract conclusions do not refer to implications of the study findings. Does this negative finding mean training is not effective (or cost effective) and should be abandoned? Or that there wasn’t the statistical power to detect a difference? Could the authors replace or add to their current text with implications of their findings.

Thank you for your comment. We have amended the abstract as requested.

Background is clear and well written.

Thank you.

Methods are not entirely clear. It appears that a group discussion was part of the intervention where barriers to adherence to guidance were discussed. How was the information collected in these groups used as part of the intervention? It sounds as if it was part of development work to
design the intervention but the inclusion of this text within the section describing the intervention implies it is part of the intervention. Could the authors clarify?

Thank you for your comment. We have clarified that this was part of intervention development (P6).

More detail of the intervention would be helpful. For example how long did it take, was it didactic or interactive, was it supplemented by written material etc? TIDIER guidance on description of interventions in trials would be a useful framework for the description of the intervention.

Thank you for your comment. We have added detail of this to the paper (P6).

The authors chose teams to participate based on whether it was felt that they did not adhere to guidelines. This could lead to selection bias and would ideally have been done through random selection or inclusion of all teams or based on data that evidenced which teams were not adhering to guidance. It seems unlikely that any team completely adhered to the guidance. Could the authors either give more detail on why it was felt that these were under-performing teams and highlight the limitation of this aspect of the methods in the discussion.

Thank you for your comment. In fact all the potential teams were included and we have added this detail of to the paper (P7).

The number of midwives in the different teams does not appear to have been taken into account in the power calculation and is not provided in the results. It is also unclear whether the analysis took into account any intervention x time effects, the multi-levels of the clustering ie hospitals, Trusts and teams, and the lack of data from teams for some of the trial.

Thank you for your comment. The number of midwifes per team has not been included in the power calculation as it in itself though highly related to the cluster size, is not actually the cluster size. The size of the midwifery team is the number of eligible births in each team. We did not examine time by treatment interactions as the study was underpowered for this comparison and this analysis had not been pre-specified. We have added this to the paper (P9).

We have added more detail to regarding the size of the midwifery teams to the results (P10).

I am not clear whether the primary outcome is the proportion of offers of a sweep or actual sweeps - in some part of the text the authors refer to offers and elsewhere to actual sweeps happening. But presumably offers don't always result in sweeps so it would be helpful to clarify this.

Thank you for your comment. The primary outcome was the proportion of women being offered and accepting membrane sweeping and we have checked the paper that that is consistent. In Table 2 we report those offered sweeping but who declined which was small (5-7%) and we have added this detail to the paper (P10).
There is an additional primary outcome of number of sweeps but the power calculation does not appear to take into account the use of two primary outcomes.

Thank you for your comment. We did not allow for any multiplicity of outcomes in our power calculation. Whilst the primary outcome, sweeping, has been reported in two different ways (number of sweeps and proportion of women swept) these two outcomes are very highly correlated and any multiplicity correction would be highly conservative. We have added a note on this in the discussion (P13).

There is no description of the measure used for knowledge before and after training in the methods section but a questionnaire is referred to in the results section. Please include details including how it was validated, and how it is scored as it is difficult to understand the results without.

Thank you for your comment. The midwives completed a Likert Scale contained within the questionnaire and the questions are detailed in Table 3.

Results.

Please provide details on number of midwives in each team rather than stating that they included an average of 10 midwives (nb. which is not consistent with the data included in figure 1) A bit more detail is given in the discussion but fuller descriptions across sites and range of sizes of teams would be helpful.

Thank you for your comment. We have added detail regarding the teams to the paper (P 10).

Discussion. This currently includes data of relevance which would be better placed in the results. For example team size and attendance rates.

Thank you for your comment. We have amended the paper (P10).

I agree with the author that use of more systematic methods with development work derived from a theoretical framework would have been more robust.

Thank you for your comment.

Reviewer #2: I read the trial paper "Evaluation of a bespoke training to increase uptake by midwifery teams of NICE Guidance for membrane sweeping to reduce induction of labour: a stepped wedge cluster randomised design", and I think the design is very interesting but, unfortunately, I don't feel comfortable enough with your statistical results to recommend publication in Trials.

Thank you for your comment. We hope with the amendments now included you will feel it is suitable for publication in Trials.
Major consideration.

Your adjusted analysis provides extreme values too far from direct unadjusted estimates. For your main outcome, you report proportions of 0.444 and 0.468, leading to a ratio of 0.95, far from you adjusted estimate of 0.90. Due to non-collapsibility, generalized linear models provide values for the adjusted estimators of the unit-level effect that are higher to the unadjusted population averaged estimation. I am not an expert in this random Poisson modeling, but my 'feeling' is that the differences between both estimates are too big, in special for, P11 L13, the average number of sweeps per woman, where an unadjusted ratio of 0.95 (0.660 over 0.701) becomes a significant adjusted ratio of 0.71. How can we merge adjusted unit effects of 0.71 to become a pooled unadjusted population effect of 0.95? If I interpret correctly, your results table 2 provides different non-significant estimates for this variable (unadjusted 0.627/0.603=0.96 vs adjusted 0.83). On this table, emergency CS shows an almost complete tie (13.2 vs 13.1) but an adjusted ratio of 0.89.

Thank you for your comment. The reviewer is correct to note that there is a discrepancy between the ratio of the two proportions and the relative risk presented in the table (RR=0.9). This discrepancy arises because the relative risk presented in the table is adjusted for time effects. In fact it is also important to note that the ratio of the percentages swept without adjusting for time is not 0.9, but is in fact 1.05 (=intervention percentage/control percentage=46.8/44.4). That is to say, the raw results suggest that on implementation the point estimate of the percentage of women being swept increased (from 44.8% to 46.8%), hence an increased "risk" of being swept. However, in actual fact, in those clusters and time periods yet to be exposed to the intervention there was an underlying secular trend. However, after adjusting for the underlying secular trend, we demonstrate a reversal of the treatment effect. Although of note, all these changes are small and not statistically significant. More discussion on this point has been included (P11 and P13).

Possible explanations are: contamination (unlikely as there was little migration between teams and intervention was only available within the trial); rising tide (I believe this is the most likely explanation – Chen YF, Hemming K, Stevens AJ, Lilford RJ. Secular trends and evaluation of complex interventions: the rising tide phenomenon. BMJ Qual Saf. 2016 May;25(5):303-10. Doi: 10.1136/bmjqs-2015-004372. PubMed PMID: 26442789; PubMed Central PMCID: PMC4853562. ) or chance finding, since none of these results are statistically significant.

If those results are correct, I wonder if such avour ng can provide useful and interpretable results. [In that case, I would suggest a more technical publication directed to one statistical journal.] I’m more in avour of some programming or copying error (despite the replication with different software by different authors). In any case, I’ll be happy to listen an explication for those discrepancies. In that case, please be also prepare to help readers to interpret such results.

Please see explanation above.

Please, discuss the discordance between the levels of instrumental delivery (the true desired outcome) and sweep, both major in the treated arm. Is this results against the NICE background to recommend sweep?
Thank you for your comment but the differences were neither statistically or clinically significant.

Other considerations

Please, extend your Consort table to include the items from the extension to cluster designs. And adapt your paper accordingly. For example, adapt table 1 to include the cluster data, adapting it to the temporal, paired nature of the SWD.

Thank you for your comment- we have done this but did not to include the number of clusters in the baseline table. The number of clusters is clearly shown in the CONSORT flow diagram. There were 10 clusters in this trial, and all 10 clusters contributed to the "intervention" and "control" data. If we were to add the number of clusters into Table 2, this would risk giving the impression that there were 20 clusters in this trial (i.e. 10 in control and 10 in intervention). For this reason, we think it is preferable not to include the number of clusters in table 2.

Please, report proportions following Consort advice. (i.e., X out of Y, Z%).

Thank you for your comment- we have done this.

Please, define what you exactly mean by pseudo-anonymized.

Thank you for your comment. Following permission from the Research and Development Departments in both Trusts data was extracted onto a study specific form using a unique identifier to enable comparisons to be made. We have added this detail to the paper (P 8).

Please, clarify in sample size and analysis if your alpha level is two-sided.

Thank you – we have clarified this in the paper (P9)