# Response to comments for manuscript entitled "Comparative efficacy of hospital hand hygiene promotion interventions: a systematic review and network meta-analysis."

**Manuscript ID: BMJ.2014.024450**

**Committee's Comments to Author:**

"Please also respond to these additional comments by the committee"

**Committee: Comment 1**

1. We tend to prioritise for publication papers that include clinical outcomes. Even if your paper doesn't do this, could you provide more of a discussion of available evidence on links between hand washing, strategies examined here, and clinical outcomes.

**Response**

We have added a section summarising key clinical outcomes at the end of the results. We now also consider this briefly in the discussion.

**Committee: Comment 2**

2. You excluded non-English studies, unpublished, and any that don't match specific quality criteria. Does this selective approach mean you are looking at a subset of available studies, not the full picture?

**Response**

Yes. We chose to exclude studies of very low methodological quality because such studies are at greater risk of bias and could introduce systematic distortion into the results. There are a lot of low quality studies in the healthcare epidemiology literature and in a previous systematic review by one of us (Cooper et al, BMJ. 2004 Sep 4;329(7465):533.) where we imposed no quality criteria or language restrictions we had to conclude that the vast majority of studies were so methodologically poor that they were of no use for evaluating the interventions. In the present review we do look at a subset of studies and we think there are good reasons for doing so: it is the subset most likely to provide reliable evidence.

The value of including non-English studies has been addressed by a number of reviews, and all have reached similar conclusions for conventional medicine. For example Moher et al (Health Technology Assessment 2003; Vol. 7: No. 41) report: “The present results, and those reported previously, suggest that excluding reports of RCTs in LOE [Languages other than English] from the analytical part of a systematic review is a reasonable way to conduct a review [random effects model (RE) ROR = 1.02; 95% confidence interval (CI): = 0.83 to 1.26]. They conclude “Language restrictions do not appear to bias the estimates of a conventional intervention’s effectiveness.” Similarly, Morrison et al (PMID: 22559755) concluded “we found no evidence of a systematic bias from the use of language restrictions in systematic review-based meta-analyses in conventional medicine”. An HTA review by Egger reached the same conclusion about language restrictions (Health Technology Assessment 2003; Vol. 7: No. 1) and also observed that difficult to locate studies tended to be of lower methodological quality, concluding “The finding that trials which are difficult to locate are often of lower quality raises the worrying possibility that rather than preventing bias through extensive literature searches, bias could be introduced by including trials of low methodological quality. We believe that in situations where resources are limited, thorough quality assessments should take precedence over extensive literature searches and translation of articles”. We do accept that in some areas of medicine where there are a large number of RCTs some of which might be withheld for commercial reasons, such unpublished studies should be included where possible. In healthcare epidemiology, however, the situation is different: there are hardly any RCTs of hand hygiene interventions but a very large number of unpublished studies of much lower methodological quality; we judged that there would have been little value (and possible harm) in including these. Moreover, they would have been very hard to locate and the resource implications for the review would have become prohibitive. We therefore followed the approach suggested by Egger et al and prioritized quality assessment.

**Committee: Comment 3**

3. We did not think publication bias constituted a major problem, but we did agree with the reviewers that heterogeneity seems high and more discussion of this is warranted, especially around the decision to pool the results.

**Response**

In the first paragraph of the "Meta analysis/Data synthesis" section of results we do now highlight the large heterogeneity which seems to result from the low fidelity to intervention in the ACE wards in the Fuller study (only 50% of wards implemented the intervention and we report intention to treat results. Per protocol ORs are actually very similar to Huis et al). We also now briefly discuss heterogeneity in paragraph 6 of the discussion.

We have also extended the treatment of publication bias by using contour-enhanced funnel plots to help distinguish publication bias from other causes of asymmetry (supplementary appendix). We now also discuss the evidence for publication bias in the results (at end of first section in the results) and mention the potential for it as a limitation in the discussion.

**Reviewer(s)' Comments to Author:**

**Reviewer #1: Overall comments**

Luangasanatip and colleagues performed a meta-analysis using rigorous methods to determine which hand hygiene interventions improve compliance. Strengths of this meta-analysis include stratifying studies by study design and using the best meta-analysis method for each design. The network meta-analysis is novel for this field and overcame the problem of no head-to-head comparisons for hand hygiene interventions. However, this meta-analysis relied too strongly on a systematic literature review performed by someone else (Gould et al.) without validating the results of that systematic literature review.

**Response**

This is a good point, and we have now gone back and individually assessed each study excluded by Gould et al (and other reviews) to confirm that exclusion criteria were indeed satisfied. This has resulted in the inclusion of 6 additional studies and the manuscript has been revised accordingly.

**Major**

**Reviewer # 1: Major comment 1**

1. Abstract, please change the abstract to say that databases were searched from 2009-Feb 2014 then supplemented with studies found by other meta-analyses.

**Response**

We have changed the abstract as suggested.

**Reviewer # 1: Major comment 2**

2. Methods, search strategy: Did you search any of the literature from 1980-2009 or did you just trust that the other reviews found all of the articles that you needed? I recommend validating the searches done before 2009 by running your search criteria during a portion of that time period (2 years or so) and determining if you would have excluded all of the studies that they excluded.

**Response**

Thank you for this helpful suggestion. We have now followed the suggestion to run the full search on selected years included in previous reviews (we chose three years rather than 2: 1980, 1995 and 2009). Two reviewers independently screened titles and abstract and, where necessary, full papers. This new search found no additional studies meeting our inclusion criteria

**Reviewer # 1: Major comment 3**

3. Similarly, it is hard to believe that 31 studies met EPOC criteria in a 5 year period (2010-2014) but only 4 studies met EPOC criteria in a 29 year period (1980-2009). Did you validate your use of EPOC versus Gould's use? Please validate the use of EPOC by Gould et al by taking studies that the Gould study excluded and applying the EPOC criteria to see if you would have excluded that study as well.

**Response**

As mentioned above, we have now followed this suggestion which has resulted in the inclusion of 6 additional studies. We have revised the manuscript accordingly.

**Reviewer # 1: Major comment 4**

4. Methods, inclusion and exclusion: Include a sentence or two on what characteristics a study must have in order to meet minimal quality criteria specified by EPOC.

**Response**

We have added this information about the EPOC criteria in the methods ("Inclusion and Exclusion" section). The new text begins "Acceptable study designs were... " Please note that the EPOC criteria were revised in Dec 2013 (after we started our review) and are now slightly more restrictive. We now use the new criteria which has resulted in the exclusion of a single study (Harne-Britner (2011)).

**Reviewer # 1: Major comment 5**

5. Methods, data synthesis and statistical analysis: describe how you tested for heterogeneity between studies (e.g. $I^2$) and how you evaluated publication bias.

**Response**

We have added this information to the second paragraph in the "Data synthesis and Statistical analysis section.

**Reviewer # 1: Major comment 6**

6. Results, RCTs: Why are there 3 studies included in Figure 3? The paragraph that describes it on page 11 mentions 4 studies and 2 studies. If the Fuller study is included twice here, ACE and ITU need to be spelled out and there needs to be a description on why it is statistically valid to include the same study twice. Additionally, an I2 of 81% means significant heterogeneity. Perhaps rather than

pooling these studies, the manuscript can just contain a description of the findings of each study. This would also cut a figure which would be useful since having 8 figures is excessive. Also, why are the results of this analysis different in the abstract and the results section of the manuscript?

**Response**

In the paper by Fuller et al. (a stepped wedge CRCT in 16 ICUs and 44 acute care of the elderly [ACE] wards) results of the intervention in the ICUs and ACE wards were reported and analysed separately (in accordance with the trial protocol) and we therefore treat these as separate studies (we are not including the same data twice). We agree that there is large heterogeneity and we now highlight this, pointing out that it arises from the fact that the intervention was not implemented in half of the ACE wards and the per protocol analysis gives almost identical results to the study by Huis et al.

**Reviewer # 1: Major comment 7**

7. Although I am not familiar with network meta-analyses, some of the odds ratios shown in Table 2 seem extremely high with extremely wide credible intervals. Please mention this in the discussion and explain the reliability of these results.

**Response**

We have modified this analysis in accordance with suggestions by reviewer 3 and now include only studies containing interventions in one of the four suggested categories (see below). The main problem was caused by i) the extremely high OR in the study by Crews et al (accompanied by extremely wide CIs) which reflects the fact that reported post-intervention compliance was close to 100% while predicted post-intervention compliance in the absence of any intervention was close to 0%; ii) the fact that for many of the comparisons we had only a single study. The changes mean that the meta-analysis now excludes a number of previously included studies (including Crews et al.) and with the few new categories of intervention for each comparison we now have at least three studies.

**Minor Comments**

**Reviewer # 1: Minor comment 1**

1. Methods, search strategy: database of abstracts of review of effects (DARE) is listed twice. Also, please spell out EPOC and ACP.

**Response**

Text amended

**Reviewer # 1: Minor comment 2**

2. If space, consistently spell out CBA and CCT since these are not standard acronyms.

**Response**

Text amended. Note that we now follow revised EPOC guidelines regarding the nomenclature of these studies so CBA is a Controlled before-after study and a CCT is a non-randomised Trial.

**Reviewer # 1: Minor comment 3**

3. Appendix 8. The first 2 funnel plots do not have labeled axes (axis) and the labels on the third funnel plot look incorrect (log odds ratios are never negative).

**Response**

We have fixed these problems. Note log odds ratios will be negative if the odds ratio is less than 1.

**Reviewer(s)' Comments to Author:**

**<u>Reviewer #2:</u>**

**Reviewer #2: Overall comments**

The authors present the results of a systematic review and network meta-analysis. Their objective was to evaluate the comparative efficacy of hand hygiene improvement interventions targeted at healthcare workers and to quantitate the resources required for such interventions.

Studies were identified from 2 previous reviews and a literature search was conducted to identify additional studies published since these prior reviews. Studies were included if they had an intervention targeting healthcare worker hand hygiene in the hospital setting, measured hand hygiene via direct observation or a proxy (e.g. product consumption, electronic or video monitoring) and used one of the following study designs: RCT or cluster RCT, controlled clinical trial, interrupted time series analysis or controlled before-after study. Studies were excluded if they were not peer reviewed or were not published in English. Additionally, studies that did not meet EPOC quality inclusion criteria were excluded although these criteria were not explicitly stated.

All included studies were systematically reviewed and summarized but only 2 RCT were meta-analyzed and only a subset of the interrupted time series studies were examined in network meta-analysis.

The key findings of the study are:

1) In a meta-analysis of 2 RCT, interventions that included all 5 components of the World Health Organizations multimodal hand hygiene program (WHO-5) plus goal setting were superior to interventions using only WHO-5.

2) In a network meta-analysis, WHO-5 and WHO-5 plus (multimodal interventions that included WHO-5 elements plus additional elements such as goal setting, incentives and accountability) were superior to standard of care / no intervention.

3) All strategies demonstrated a trend towards improved hand hygiene compared with standard of care / no intervention and all WHO plus (WHO-5 with additional interventions including goal setting, incentives or accountability interventions) demonstrated a trend towards improved directly observed hand hygiene compliance compared with WHO-5 though confidence intervals were wide and overlapping.

4) Insufficient data on costs were presented in the literature to allow meaningful conclusions but some approximate ranges are presented

Assessment

Healthcare worker hand hygiene is a critical strategy to reduce the global burden of healthcare-associated infection (HAI). Given that healthcare workers adherence to current hand hygiene guidelines remains suboptimal, a comparative evaluation of interventions to improve hand hygiene compliance in healthcare is of vital importance.

This manuscript should be highly relevant to policymakers, hospital administrators and infection prevention and control programs that are actively pursuing quality improvement interventions to reduce HAI incidence and enhance patient safety. While the optimal approaches to improving healthcare worker hand hygiene compliance is of vital importance, it may be perceived as less relevant to front-line healthcare workers not directly involved in developing quality improvement efforts in this area.

I believe that this article is original and I am not aware of other systematic reviews on this topic that used network analysis. As the authors themselves highlight, there are other systematic reviews on this topic but some are now dated due to the large volume of recent publication in this area while more recent reviews made different methodological choices in terms of study selection and quality assessment and did not use network meta-analyses. For these reasons, I think this study makes an important and original contribution to the field.

The study is well written and is fairly clear although the amount of information conveyed is large and at times it is challenging to follow. I believe some improvements to clarity could be made. The study question and study design are appropriate, with the caveat that I have limited experience with network meta-analysis. The results appear valid and the conclusions follow logically from the results. I do have suggestions for revisions, which follow below.

**Response**

Thank you for these constructive comments. We believe the changes we have made in response to all reviews have helped to improve the clarity.

**Reviewer #2: Abstact comment 1**

Suggestions for Revisions

Abstract

The Design and Inclusion sections of the abstract do not adequately explain the methodology used by the study. Given that the current abstract is brief (<350 words) and assuming that a longer abstract would be acceptable, I believe more details could be provided to ensure that that abstract can be understood as a stand-alone document. For example, it does not described the data sources used, the type of study designs included or the designs and outcomes relevant to the network meta-analysis (e.g. the study included a variety of designs and outcomes, but only interrupted time series that measured directly observed hand hygiene were included in the network-meta-analysis). As such, the abstract lacks several elements suggested for inclusion in a 'structured abstract' as described by the 2009 PRISMA checklist (http://www.prisma-statement.org/2.1.2%20-%20PRISMA%202009%20Checklist.pdf).

**Response**

We have modified the abstract as suggested.

**Reviewer #2: Introduction comment**

Introduction

In the first paragraph, the references 1 and 2 supporting the statements on the burden of HAI are old. Newer primary data on this topic are available (e.g. Magill et al, NEJM 2014).

**Response**

Thank you for suggestion on the updated citation. However, the suggested study, Magill et al.2014, focuses on the US but for this paper we think it is more appropriate to focus on the global picture so would prefer to keep existing references. If there are more recent or more accurate global estimates which we are not aware of we would be happy to use them.

**Reviewer #2: Methods comment 1**

Methods

Page 5, paragraph 1: It is stated that the PRISMA statement was used to guide reporting of the study however several elements have been omitted (in addition to the structured abstract described above):

1) The objective is not stated explicitly in the introduction using the 'PICOS' format

**Response**

We have modified the abstract as suggested and now explicitly state the objective in the introduction as suggested.

**Reviewer #2: Methods comment 2**

2) A registration number for the protocol was not provided (nor a statement indicating that the protocol was not registered)

**Response**

We have added a statement to the methods (first paragraph) indicating that the protocol was not registered as suggested.

**Reviewer #2: Methods comment 3**

3) A list of variables abstracted from all studies was not provided.

These elements should be added to the manuscript if possible.

Page 6, paragraph 2: The term retrospective is used but not defined. An explicit definition should be given as this term is used variably.

**Response**

We have now provided the additional information as suggested.

**Reviewer #2: Methods comment 4**

Page 6, paragraph 3: It is stated that studies were excluded that failed to meet 'minimal quality criteria specified by the Cochrane Effectiveness Practice and Organisation of Care Group (EPOC). I was unable to find the reference linked to this statement (ref 12) although I believe I did find the information online at the Cochrane website. These criteria should be clearly outlined in the paper and a definition of what 'minumum' thresholds are should be presented - as well an appropriate reference to either a published article or to a website that includes an accurate URL should be added. In Appendix 5, the reasons for exclusion of these studies are given and are these exclusions appear to be mainly on the basis of study design (e.g. study was an uncontrolled before-after study) and I found this a bit confusing as this design did not meet the authors 3rd inclusion criteria on page 6 (study design criteria) and should have been excluded at an earlier stage in the process (i.e. at the stage of full text review at least)?

**Response**

We have added a new citation for EPOC minimum requirements for each study design and provided a better citation with URL. We now also clearly outline these criteria in the paper as suggested. In addition, we also clarify the issue about inclusion criteria and EPOC criteria – appendix 5 lists studies that met inclusion criteria but not EPOC criteria. Reasons for not meeting EPOC criteria are now summarized in figure 1.

**Reviewer #2: Results comment 1**

Results

Page 10, paragraph 5: A cost range is presented ($US 225 to $4669 per 1000 bed days). The study associated with the highest cost included only the cost of one time video camera installation. That study involved video footage outsourced to an external group that reviewed the video and estimated hand hygiene performance on room entry and exit. The human resource costs associated with reviewing the video were likely substantial so I do not think $4669 per 1000 bed days is a meaningful estimate of the costs for this intervention. I suspect there are similar issues with the cost estimates for many of these studies and wonder whether the presentation of any quantitative data here is useful at all? The authors themselves conclude that reporting of resource use is inadequate in the literature and I think this is the only meaningful finding with respect to resource utilization.

**Response**

We agree that the main message is that quantifying resources required for interventions was inadequate. As suggested, we have now taken out the range of costs summarized in the main text in the results though think it is still worth summarizing in the abstract to highlight two things: i) we have summarized information on costs/resources from these studies, so anyone interested in this should read our paper(!); ii) the range is very large.

**Reviewer #2: Results comment 2**

Page 11, paragraph 1: It is stated that 2 RCT demonstrated improved compliance following implementation of education, performance feedback and visual reminders (ref 34) or education alone (ref 32) but it is not explicitly stated what occurred in the control arm. The answer is in table 3 but it would be easier for the reader if it were stated here.

**Response**

The recommended additional information has now been added to the results section as well as more details of the findings from these two RCTs.

**Reviewer #2: Results comment 3**

Page 11, paragraph 2 discusses the results of the included randomized controlled trials. This methodology is stronger and less prone to bias than the controlled before-after or time series designs. I noted with interest that the RCT included here tended to be associated with small absolute increases in compliance. Would it be possible in your review to discuss the magnitude of improvement rather than just qualitatively whether hand hygiene improved? Is there a correlation between study quality and a lower absolute improvement in hand hygiene? What is a clinically significant increase in hand hygiene compliance given that these studies have large sample sizes and can therefore detect small differences (e.g.

the RCT by Fisher et al. observed 1,017,600 opportunities for hand hygiene using an automated detection system while the cluster RCT by Mertz et al using direct human observation still observed almost 8000 opportunities).

As stated above, I will point out that:

1. In the FIT trial (ref 31) the increase in hand hygiene associated with the intervention was only 7% to 9% and in the intention to treat analysis was seen only in the ICU setting and not in the ward setting.

2. In the cluster RCT by Mertz et al. (ref 34) hand hygiene increased only 6% and there was no difference in their infection outcome (hospital acquired MRSA colonization) suggesting either that hand hygiene is not effective in reducing MRSA colonization, or that a larger improvement in hand hygiene is required to see a benefit.

3. In the RCT by Fisher (ref 30) the abstract described a 6.8% increase in hand hygiene compliance but this result is not presented anywhere else in the paper. Much of the apparent benefit appeared to be due to a drop in compliance in the control arm.

**Response**

We do now report the magnitude of improvement in each study as suggested including that by Fisher. Directly correlating effect sizes with study designs (or other indicators of study quality) is complicated by the fact that different interventions are compared by different studies. However, as study quality generally does increase with study size the funnel plots in the appendix do capture this information. Moreover, the revised funnel plots we have used do, in theory, help to distinguish publication bias from other sorts of bias. In most cases, however, there are too few studies considering the same interventions to allow meaningful comparisons. There is one important exception.

In both the FIT and Mertz studies WHO-5 was compared with WHO-5+. While none of the ITS studies directly made this comparison the network analysis allows an indirect comparison and gives results broadly in line with this though with large uncertainty (mean odds ratio for WHO-5+ v WHO-5 is 1.8 (95% CI 0.2

to 12.2). We also note that the low effect size in elderly wards in Fuller et al seems to have been because only half of elderly the wards implemented the intervention. We have now added this information to the results section.

**Reviewer #2: Results comment 4**

Page 11, paragraph 4: The authors note that in 11 of 19 comparisons among the time series analyses, hand hygiene was falling prior to the intervention. Does this suggest a bias or regression to the mean phenomenon? Was hand hygiene rising in the other 7 or was it stable? Does this suggest that the 'control' arm or 'baseline' arm was not actively engaged in hand hygiene improvement and does that complicate your classification of the control arm hand hygiene promotion strategy in that the strategy may no longer have been 'active'?

**Response**

We think regression to the mean effects would be an important issue if interventions were made when hand hygiene compliance happened, by chance, to be particularly low. In that case even an entirely ineffective intervention might be expected to show an effect as a result of regression to the mean. Having a long baseline period does protect against this (because over longer time periods such fluctuations would tend to even out) and we note that most ITS studies do have at least 3 months baseline data. Moreover, amongst those that have shorter baseline periods (Dubbert, Tibbals, Khatib, Armelino) hand hygiene was increasing in the baseline in 3 out of 4 cases suggesting regression to the mean effects are not important. Overall, amongst included ITS studies, we now find hand hygiene is increasing or stationary in the baseline period in 10 and decreasing in 12. Overall, therefore, it doesn't look like regression to the mean is a major concern here (a sharp contrast the healthcare epidemiological literature with clinical outcomes where regression to the mean effects are a major problem). It is not clear to us that hand hygiene rates shouldn't fall over time even with "active" interventions though we do note that in the 12 studies where pre-intervention hand hygiene was falling no hand hygiene promotion activities were reported in 7 studies and in another 4 there was only either an educational intervention or reminders in the baseline period. There was only a single case when hand hygiene compliance was falling when a multi-modal intervention was in place at baseline (Talbot et al) and even that rate of decline was very low. We have now added half a sentence to the results (second paragraph of ITS section) to highlight the fact that in most cases where pre-intervention declines were occurring they were in the absence of multimodal interventions.

**Reviewer #2: Results comment 5**

Page 12, paragraph 2: It is stated that Mayer et al [ref 59] used an 'appropriate' analysis. This term is vague and is not used consistently in the paper to describe other studies.

**Response**

We agree with the reviewer so this term has been deleted.

**Reviewer #2: Results comment 6**

Page 13, paragraph 2: it is stated that "…all intervention strategies were associated with an improvement in hand hygiene compliance compared with T1". However, for interventions T2, T3 and T6 in the figure appear to have confidence intervals that cross over with T1. Perhaps this statement should be softened to indicate that there was a trend to benefit for all interventions? In the network meta-analytic framework, what is the criteria to define an intervention 'associated with' improvement and can this be stated explicitly in the methods?

**Response**

This was based on point estimates. It is still possible to have an association that doesn't reach statistical significance at a particular level (and for any association there will be some significance level where significance is not reached). Sensible recommendations are usually not to arbitrarily dichotomize into significant/non-significant at some arbitrary level (eg. http://www.bmj.com/content/322/7280/226.1) but to consider confidence intervals. We have tried to follow this advice and base interpretation on point estimates and confidence/credible intervals (and we haven't applied any decision rules based on whether OR intervals include 1 or not). However, given the uncertainty, we do agree that the statement probably needed softening a little. Also, because of the regrouping of strategies (following reviewer 3's suggestions) we have a lot more clarity in the network analysis results. What we now say is “The network meta-analysis showed that although there was large uncertainty in effect sizes amongst the pairwise comparisons, point estimates for all intervention strategies indicated an improvement in hand hygiene compliance compared with no intervention (Figure 7). For two strategies, WHO-5 and WHO-5+, when compared with no intervention there was strong evidence that they were effective (Table 2).”

**Reviewer #2: Results comment 7**

Page 13: A system of naming ITS interventions as T1 through T12 is introduced. Despite the table that explains these terms, it is hard to keep track of the correlation between the name and the intervention. It would be better to use abbreviations that captures the nature of the intervention itself (e.g. WHO-5+I could be used to indicate a study using the 5 WHO interventions plus incentives).

**Response**

We agree. This was confusing. We have now reclassified the categories of different intervention strategies (following reviewer 3’s suggestion), and we think the new system should be easier to understand. We have also tried to avoid uninformative labels (T1, T2) throughout the revised manuscript and have used abbreviations as suggested instead.

**Reviewer #2: Discussion comment 1**

Discussion

Page 15: The limitations of the study are well described. Two additional limitations that perhaps should be discussed include:

1. Because the network analysis compares a new intervention with a baseline intervention is this a limitation because more energy and attention may be directed to the new intervention while the control intervention is ‘old news’? is the benefit seen with almost any intervention described here (compared to baseline) simply a reflection that a new intervention will lead to a transient improvement in hand hygiene?
2. The limitations of direct observation as a means of recording hand hygiene are not discussed, particularly with regards to observation bias (Hawthorne Effect) and how it might complicate the interpretation of this data.

**Response**

We have added both of these points as limitations in the discussion.

**Reviewer #2: Discussion comment 2**

Page 15, paragraph 2: It states that there is no asymmetry in the funnel plot. To my eye the plot does look somewhat asymmetric. Is there a more objective measure that can be used to determine if asymmetry or potential publication bias was present? A priori, one might expect that this is a field where publication bias is quite likely to occur.

**Response**

We have now used improved funnel plots which aim to help distinguish between possible causes of funnel plot asymmetry (the enhanced contour funnel plot). Eyeballing the enhanced contour funnel plots we don't see any clear evidence of asymmetry leading to fewer studies in the $p>0.05$ region (which would suggest publication bias), though certainly publication bias seems quite plausible *a priori*. However, given that there were no more than four pair-wise comparisons for any pair of strategies we don't think there are sufficient data to say anything definitive about this except that it is a possibility. We have added some text at the end of the first section of the results to make this clear.

**Reviewer #2: Conclusions comment 1**

Conclusions

The conclusions are clear and follow logically from the results.

Figures and Tables

Figure 1 should provide the rationale for excluding studies to get from 136 to 36 studies (this is provided in appendix 5 but could be easily summarized here as there were a limited number of reasons for exclusion).

**Response**

A summary of reasons for exclusion has now been added to Figure 1.

**Reviewer(s)' Comments to Author:**

**<u>Reviewer #3:</u>**

**Reviewer #3: Overall comments**

This Systematic review evaluates the evidence for hospital hygiene promotion with particular emphasis on that identified for WHO-5 (System change, Education, Feedback, Reminders, and institutional safety climate). Reduced evidence was identified from RCTs but the inclusion of other study designs, in particular Interrupted Time Series, provides more information about the comparative effectiveness of several strategies. There is a major issue in that the current use of Network Meta-analysis given the sparse data is not adequate. This is highlighted by the estimate obtained for T8 (obtained from a double indirect comparison: T8 vs T3, T3 vs T11, T11 vs T1).

One potential solution is to drop altogether the NMA and focus only on the direct pairwise comparisons while an alternative is to restructure the current network and reclassify the interventions into 4 categories:T1, T2 to T6, T7, T7+. This way, a complete network could be created (leaving nodes not in the network out, such as T4, T5, and T8) and a comparison of WHO5 vs. None, WHO-5 vs WHO-5+ could be made. From the Introduction and the Discussion, this appears to be the main focus of the paper. The choice will clearly depend on the coherence of this grouping. If not appropriate, please drop the attempt to use NMA and just report pairwise comparisons. Clearly this change will mean a readjustment of the rest of the paper.

Please also provide a reference for the analysis of the ITS presented in Appendix3.

**Response:**

Thank you for this very helpful suggestion. We have made these suggested changes (reclassifying interventions as suggested) and think this has resulted to a major improvement to the manuscript.

We have also added a reference to the segmented regression analysis for binomial outcomes (Taljaard et al. 2014, *Implementation Science* 2014, **9**:77 doi:10.1186/1748-5908-9-77).