PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Compliance with a Time-out Procedure intended to Prevent Wrong Surgery in Hospitals: Results of a National Patient Safety Program in the Netherlands
AUTHORS	Van Schoten, Steffie; Kop, Veerle; De Blok, Carolien; Spreeuwenberg, Peter; Groenewegen, Peter; Wagner, Cordula

VERSION 1 - REVIEW

REVIEWER	Atsuhiko Murata Department of Preventive Medicine and Community Health, School of Medicine, University of Occupational and Environmental Health
REVIEW RETURNED	22-Mar-2014

GENERAL COMMENTS	I think that this paper is well studied and well written. However, some revisions are required. Please consider the following points. 1. Authors investigated the compliance with the time-out procedure (TOP) according to the kinds of surgery (general, gynecologic or urologic surgery). However, authors did not survey the other aspect of surgery such as elective or urgent surgeries. Authors had better add the data about comparison of the compliance with TOP between elective or urgent surgeries. 2. Authors should explain the definition of academic and teaching hospitals. What is the difference between academic and teaching hospitals?

REVIEWER	Paulo Sousa National School of Public Health, Portugal
REVIEW RETURNED	01-Apr-2014

GENERAL COMMENTS	Main comments: - In my opinion the title goes beyond the contents of the manuscript. The authors doesn't have any data/results about "prevent wrong surgery in hospitals", as they mention in the title. I suggest they could change the title. - It is not clearly how the authors measured "completeness"? Is there any "guideline" or norm for that? - It is widely known the "Hawthorne effect" when Observation is used as a technique of measure. Probably the authors can say anything (discuss the potential limitation or bias effect in this study) about this in the discussion part?

	Minor comments: - Can you explain the rate of 4.8% of unknown (related to gender)? - What is the meaning of ENT page 10 and 11? - The authors didn't mention how much observers participated in the study (variability inter observers). - "OR" appear in page 4 and only in page 6 "operating room" (OR). - In the discussion part it would be interesting if the authors could "talk more" (discuss) about the results they found and the possible implications in the implementation and success of the "national safety program", or for further studies. This is just a suggestion, probably it goes beyond the aim of the manuscript....

REVIEWER	Penny Rhodes and David Springate Centre for Primary Care Research, University of Manchester, UK
REVIEW RETURNED	21-Apr-2014

GENERAL COMMENTS	The stated aims of this paper were 'to evaluate the extent to which hospitals carry out the TOP before anesthesia in the operating room, whether compliance has changed over time, and to determine factors that are associated with compliance.' However the authors do not justify why the specific factors investigated were selected and there appeared to be no overt theoretical underpinning or hypothesis testing. I am not convinced that type or size of hospital are particularly useful potential explanatory factors. I would also question why type of hospital was used in the analysis when the number of academic hospitals (n=2) was always going to be too low for any meaningful conclusions? Was one of these the hospital 'that never performed the TOP correctly' (p5, lines 32-3)? Perhaps, the results could have been presented both with and without that particular hospital included. Evidence that TOP leads to reduction in wrong site, procedure, person errors is limited and this needs to be highlighted in the introduction/discussion. Moreover, equating TOP compliance with more/less safe practice is also somewhat simplistic; one would need to look at performance of the procedure as a whole and to examine on what grounds clinicians were making decisions about non-compliance, before concluding that it was less safe. The more interesting findings – that compliance is lower with older patients, and that team membership was often incomplete and/or members were often not fully focussed on the TOP, and that there was wide variation between individual hospitals – could have been discussed in greater depth. Specialist statistical review Given the results, I think it is questionable to focus on the differences between the hospital types. Although the descriptive stats show a large difference in mean TOP scores between the Academic hospitals and the others, table 1 shows that in every case but Measuring moment 5, the confidence intervals overlap between all of the hospital types and the mean trend. What is clear is that the variance in TOP scores for academic hospitals is far greater than in the other hospital types. What the paper doesn't seem to provide is a breakdown of numbers of procedure by hospital type, so it is impossible to tell if this increased variance is simply because of a really low number of procedures.

	The multilevel analysis is the correct way to analyse this data, but I am also concerned that the different measurement points (presumably the same as time points and measuring moments? The paper is a little unclear on this) were pooled after the first analysis. They would not be independent of hospital and should be modelled as such. Pooling the timepoints suggests that the analysis is suffering from pseudo-replication. It seems curious that the authors did not do the multilevel modelling for the effect of hospital type on the TOP score in the pooled analysis, like they did with specialities, checking and focus, or include this in the models for them. Could they comment on why this was? I guess it is because there were not enough units at this level but it would be useful to have a breakdown of the numbers of procedures in each group. It is interesting that the patterns of peaks and troughs in the time series looks similar for the different hospital groups. Could the authors comment on why this may be? Finally, how valid is TOP compliance as a proxy for wrong surgery? For example is there any suggestion in the literature that Academic hospitals have poorer safety records etc.? In summary, though the authors recognise in the discussion, these finding cannot be generalised to academic hospitals as a whole, that academic hospitals have lower compliance is one of the major reported findings. I think that this claim should be moderated and not made to be the main focus of the paper. As a minimum, the authors should present a full breakdown of procedure numbers by hospital type, specialities, checking and focus. The raw data should be presented as a table broken down by hospital (grouped within type), time points, procedure and %TOP compliance.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1
1 Authors investigated the compliance with the time-out procedure (TOP) according to the kinds of surgery (general, gynecologic or urologic surgery). However, authors did not survey the other aspect of surgery such as elective or urgent surgeries. Authors had better add the data about comparison of the compliance with TOP between elective or urgent surgeries.

Response authors:
Thank you for your suggestion, we appreciate that it would be indeed interesting to compare compliance with the TOP between elective and urgent surgeries. However, data on urgent surgeries was not available as we only observed elective surgeries. This was decided for practical as well as safety issues, as the team needed to be informed about the presence of the observer in the OR and there would be no time to inform the team in urgent situations.

2 Authors should explain the definition of academic and teaching hospitals. What is the difference between academic and teaching hospitals?

Response authors:
In the Netherlands, teaching hospitals provide specialized medical care and are committed to training and education. The level of care can be characterized as complex and lies between that of general hospitals and academic centers.

We added a description of the difference between academic and teaching hospitals to the methods section. See page 9.

Reviewer 2
1 In my opinion the title goes beyond the contents of the manuscript. The authors doesn't have any data/results about "prevent wrong surgery in hospitals", as they mention in the title. I suggest they could change the title.
Response authors:
In the title, the part 'to prevent wrong surgery' refers to the TOP and not to the results of our study. We do not intend to claim in the title that we have data on the effect of the TOP on the actual incidence of wrong surgery. However to avoid confusion we added ‘intended’ to the title:
‘Compliance with a Time-out Procedure intended to Prevent Wrong Surgery in Hospitals: Results of a National Patient Safety Program’

2 It is not clearly how the authors measured "completeness"? Is there any "guideline" or norm for that?

Response authors:
The complete team in this study was seen as the group of persons that performed the surgery on the patient. To be able to perform a TOP correctly, the complete team was present during the TOP. When this was not the case, meaning that one or more persons joint the team after the TOP had been completed, team completeness was scores as ‘no’.
We added this explanation in the methods section on page 9.

3 It is widely known the "Hawthorne effect" when Observation is used as a technique of measure. Probably the authors can say anything (discuss the potential limitation or bias effect in this study) about this in the discussion part?

Response authors:
Thank you for pointing this out. We mention the potential observation bias already in the discussion section under strengths and limitations, and also the method that we have used to minimize this effect. We think this suffices.

4 Can you explain the rate of 4.8% of unknown (related to gender)?

Response authors:
In one of the participating hospitals it was not allowed to collect data on the gender of the patient, as they wanted to ensure complete anonymity. Therefore, we have no information on gender in 4,8% of the observations. Given the relatively small percentages of the total of observations we don't expect an influence on the results as presented.

In the manuscript we have changed ‘4,8% unknown’ in ‘4,8% not registered’.

5 What is the meaning of ENT page 10 and 11?

Response authors:
ENT is an abbreviation for Ear, Nose and Throat Medicine.
We added this explanation to the text on page 12.

6 The authors didn't mention how much observers participated in the study (variability inter observers).

Response authors:
Twelve observers participated in this study. To limit inter observer variability, all observers were trained prior to the start of the observations. Moreover and regular feedback meetings were held where observers exchanged experiences and discussed how to deal with certain situations and observations at the OR.

Furthermore, standardized registration forms were used.

We added the training and feedback meetings to the methods section on page 8; the standardized registration form was already mentioned in the manuscript.

7 "OR" appear in page 4 and only in page 6 "operating room" (OR).

Response authors:
We corrected the use of the abbreviation according to your comment: we now write out the term 'operating room' in page 4 were it first appears and consistently use the abbreviation 'OR' from there on.

8 In the discussion part it would be interesting if the authors could "talk more" (discuss) about the results they found and the possible implications in the implementation and success of the "national safety program", or for further studies. This is just a suggestion, probably it goes beyond the aim of the manuscript....

Response authors:
Thank you for your suggestion. We acknowledge the fact that it would be interesting to discuss the implementation of the Safety Program but agree with the reviewer that it goes beyond the aim of this manuscript as the TOP was only one part of the whole Safety Program.

Reviewer 3
1 The stated aims of this paper were
'to evaluate the extent to which hospitals carry out the TOP before anesthesia in the operating room, whether compliance has changed over time, and to determine factors that are associated with compliance.'
However the authors do not justify why the specific factors investigated were selected and there appeared to be no overt theoretical underpinning or hypothesis testing. I am not convinced that type or size of hospital are particularly useful potential explanatory factors.
I would also question why type of hospital was used in the analysis when the number of academic hospitals (n=2) was always going to be too low for any meaningful conclusions? Was one of these the hospital 'that never performed the TOP correctly' (p5, lines 32-3)? Perhaps, the results could have been presented both with and without that particular hospital included.

Response authors:
Our research is explorative and the results are based on the data that was available from the Safety Program and limited to the variables that were included in the design of the program. Therefore we didn't include a theoretical framework in the introduction. However, in the discussion we reflect on the factors that we included in this study and their possible explanatory value in the light of existing literature.

Our sample of hospitals is a representative reflection of the Dutch hospital population, there are only 8 academic hospitals in the Netherlands and we have therefore chosen to present the results for this group as well.

2 Evidence that TOP leads to reduction in wrong site, procedure, person errors is limited and this needs to be highlighted in the introduction/discussion. Moreover, equating TOP compliance with more/less safe practice is also somewhat simplistic; one would need to look at performance of the procedure as a whole and to examine on what grounds clinicians were making decisions about non-compliance, before concluding that it was less safe.

Response authors:
Thank you for pointing this out. Authors agree that evidence that the TOP leads to a reduction of wrong surgery is limited, and highlighted this more in the introduction. See page 6.

We also added to the discussion section that readers should bear in mind that there are other aspects to be considered before concluding that surgery is less safe without a TOP. See page 15.

3 The more interesting findings – that compliance is lower with older patients, and that team membership was often incomplete and/or members were often not fully focussed on the TOP, and that there was wide variation between individual hospitals – could have been discussed in greater depth.

Response authors:
Thank you for your comment; we added reflections and possible explanations for our finding to the discussion part of our manuscript. See page 14 and 15.

Reviewer 4 Specialist statistical review
1 Given the results, I think it is questionable to focus on the differences between the hospital types. Although the descriptive stats show a large difference in mean TOP scores between the Academic hospitals and the others, table 1 shows that in every case but Measuring moment 5, the confidence intervals overlap between all of the hospital types and the mean trend. What is clear is that the variance in TOP scores for academic hospitals is far greater than in the other hospital types. What the paper doesn't seem to provide is a breakdown of numbers of procedure by hospital type, so it is impossible to tell if this increased variance is simply because of a really low number of procedures.

Response authors:
The authors agree that, given the low number of academic hospitals that participated in our study, it is unjust to put the main focus of the manuscript on the poorer performance of academic hospitals. However, in the Discussion part of the manuscript the authors emphasize that these results should be interpreted with caution and should not be generalized to the population of academic hospitals. In response to the reviewer, we adjusted the Conclusion part of the manuscript, to ensure that readers do not generalize these findings to academic hospitals in general. See page 15.

Furthermore, the authors added the raw data for the observations, broken down by hospital type and measuring moment to Table 1. Raw data for the remaining variables (specialty, focus and individual checks) is available upon request.

2 The multilevel analysis is the correct way to analyse this data, but I am also concerned that the different measurement points (presumably the same as time points and measuring moments? The paper is a little unclear on this) were pooled after the first analysis. They would not be independent of hospital and should be modelled as such. Pooling the timepoints suggests that the analysis is suffering from pseudo-replication.

Response authors:
Reviewer is correct that the terms 'measurement points', 'time points' and 'measuring moments' refer to the same concepts in our analysis. For reasons of uniformity and clarity we changed the terms

'measurement points' and 'time points' to 'measurement moments' throughout the manuscript.
The trend analysis showed that there was no linear trend in performance of TOP and therefore the observations for the different measurement moments were pooled in the remaining analyses. The authors agree with the fact that the measurement moments are not independent of hospitals and accounted for this by modeling a two-level structure with hospitals at the highest level and observations at the lowest level of analysis.

3 It seems curious that the authors did not do the multilevel modelling for the effect of hospital type on the TOP score in the pooled analysis, like they did with specialties, checking and focus, or include this in the models for them. Could they comment on why this was? I guess it is because there were not enough units at this level but it would be useful to have a breakdown of the numbers of procedures in each group.

Response authors:
There were not enough units at the highest level (hospitals) to model the effect of hospital type on the TOP score in the pooled analyses. The breakdown of the number of observations for each hospital type that we added to Table 1 should make this point clear to the reader. To be coherent, we've also added a sentence to the Statistical analyses part of our manuscript that explains why we couldn't model the effect of hospital type on the TOP score in the pooled analyses. See page 10.

4 It is interesting that the patterns of peaks and troughs in the time series looks similar for the different hospital groups. Could the authors comment on why this may be?

Response authors:
Authors do not agree that the patterns in the time series are very similar for the different hospital groups. The peaks and troughs are not consistently the same for measurement moments 1-8 for the different hospital types. The pattern is only consistently the same for measurement moment 8 and 9. At first we sought the explanation for the trough in measurement moment 9 in public holidays (shortness of staff and different team compositions due to holidays) but a further look in the data showed that this was not the case as the measurements were all outside holiday periods.

5 Finally, how valid is TOP compliance as a proxy for wrong surgery? For example is there any suggestion in the literature that Academic hospitals have poorer safety records etc.?

Response authors:
See response point 2, Reviewer 3.

6 In summary, though the authors recognise in the discussion, these finding cannot be generalised to academic hospitals as a whole, that academic hospitals have lower compliance is one of the major reported findings. I think that this claim should be moderated and not made to be the main focus of the paper. As a minimum, the authors should present a full breakdown of procedure numbers by hospital type, specialities, checking and focus. The raw data should be presented as a table broken down by hospital (grouped within type), time points, procedure and %TOP compliance.

Response authors:
See response point 1, Reviewer 4.