PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Strengths and weaknesses of working with the Global Trigger Tool method for retrospective record review - Focus group interviews with team members
AUTHORS	Schildmeijer, Kristina; Nilsson, Lena; Perk, Joep; Årestedt, Kristofer; Nilsson, Gunilla

VERSION 1 - REVIEW

REVIEWER	Jacob Anhøj Consultant, MD, DIT Danish Society for Patient Safety Denmark No competing interests
REVIEW RETURNED	08-May-2013

THE STUDY	The research question is defined as "describe experiences of working with GTT". Reading the article one gets the impression that the objective is to learn from experienced GTT teams with the intent of improving the GTT method. Please clearify the aim. The interviewer who did the focus group interviews should be described in more detail. Especially one needs to know more about the interviewers views on GTT before the study. The manuscript would benefit from a review from a native English speaking person.
RESULTS & CONCLUSIONS	The message is clear, that GTT should be considered a useful qualitative method, and that GTT is of less use as a quantitative method for measuring changes in patient harm over time. I do not think that this conclusion is supported by the study findings. The study design does not allow for strong conclusions on the quantitative properties of GTT. Furthermore, the study finds that the review teams had modified the review methodology significantly over time. This invalidates any conclusions with regards to the "original" GTT methodology.
GENERAL COMMENTS	Since my colleague, Brian Bjørn whom I have discussed this review with, is also reviewing this manuscript, I suggest that any revision of the manuscript is send to a thirds reviewer.

REVIEWER	Ellen Tveter Deilkås MD PhD Senior Scientist Akershus University Hospital Norway
REVIEW RETURNED	15-May-2013

THE STUDY	The objective of the study should be more specific about the aims for describing the experiences. The context for how the teams have participated in the study is not made clear. At what level in their organizations have they performed the GTT and with what mandate? How have the teams been trained and followed up, to ensure that they perform according to the original protocol and definition? Such information is relevant for understanding how the teams experiences are relevant to other GTT teams. The research questions could be presented more explicitly in the paper. The study lacks questions related to the how the core measurements of the GTT protocol have been analyzed and presented. GTT is tailored for analyzing the data with statistical process control (SPC). What are the teams experiences with interpreting the data using SPC?
RESULTS & CONCLUSIONS	There is a weak relation between the themes which the paper highlights as raised by the teams and the themes highlighted in the papers conclusion.
GENERAL COMMENTS	The context for how the teams have performed the GTT, is not made clear. At what level in their organizations have they performed the method and with what mandate? How have the teams been trained and followed up, to ensure that they perform according to the original protocol and definition? Such information is relevant for understanding how the teams experiences are relevant to other GTT teams. The Swedish GTT protocols differs from the original in that it considers the preventability of AE's. This element has raised points of concern amongst the teams which do not apply to the original GTT protocol. The article should differentiate clearly which experiences that apply to the GTT protocol, and which experiences that apply to the modified version. Some of the teams experiences have been made within contexts that differ from the purpose of the original GTT protocol. The relevance of these experiences should be specified separately from those that have been made when sticking to the original protocol and purpose. The study lacks questions related to the how the core measurements of the GTT protocol have been analyzed and presented. GTT is tailored for analyzing the data with statistical process control (SPC). What are the teams experiences with interpreting the data using SPC?

	The research questions could be presented more explicitly in the paper. There is a weak relation between the themes which the paper highlights as raised by the teams and the themes highlighted in the papers conclusion. **Page 2** Line 5 Please be more specific about the aims of describing the experiences. Line 25 The teams have evaluated their experiences with the GTT, and they have not evaluated the strengths and weaknesses of the GTT according to the protocols aims and standards. There is a difference here which the authors should be clear about. To evaluate the GTT according to its aims and standards would have required a more standardizing approach, than which is applied in this study. Line 38-45 The first sentence does not specify which difference it refers to, although one can guess that it is the teams, it should be specified. The second sentence lacks logical foundation in the paper. **Page 3** Line 30 This sentence lacks logical foundation in the paper and should be revised as suggested elsewhere in this review. Line 40-41 This sentence lacks foundation and reasoning in the paper.

	Page 6 Line 23-25 Please clarify. In this sentence it seems like finding triggers is an aim, more than a mean, to find the AE's. **Page 11** Lines 10- 21. It is not clear if the nurses registered the findings as AE's according to the GTT, even if the findings did not comply with the AE definition, or if they registered the findings in other ways. Please clarify. **Page 12** Lines 36-38 Please explain how the comment is relevant to the GTT definition of harm, which only relates to physical harm. Explain in which other ways this comment might be relevant. **Page 13** Line 10-14 Please explain how the comment is relevant to the GTT protocol **Page 14** Line 7-14 It seems if the tool has been used to facilitate dialogue related to the single events rather than to make measurements, based on statistical process control, as GTT is tailored for in the original protocol.

	Page 15: Line 45-47 needs clarification. **Page 16** Line 14 Please specify for what and how, the teams found the GTT method useful. Line 30-40 It is not clear if the teams report that they have changed over time, how they supposedly perform the GTT according to the protocol, or if they have chosen to use a deliberately modified version. A clarification of this is important for understanding the consistency of the method and how it is conditioned. Line 39 Referring to a team that obviously does not follow the protocol, may reflect differences in teams capability to apply the protocol. This should probably also be discussed. Line 43 Is a punctuation mark lacking? Line 43 -48 It should be emphasized that the adjustments which Kennerly describes, were made before the method was implemented so that the review could be conducted consistently over time. This facilitated a good interrater reliability amongst reviewers (K = 0.62). Line 54 -56 The reasoning behind the conclusion lacks some components.

	1. Changes in methodology. The paper says little about on what conditions the teams have changed their practice of the methodology. 2. Assessment of preventability. This element does not apply to the original GTT protocol. Conclusions on how the original GTT protocol should be used can not be based on experiences with an element which it does not contain. **Page 17** Line 3-4 It is possible to suggest that the GTT method can produce information on AE's that is useful in qualitative enquiry. But that does not exclude it's original purpose, as the authors have concluded. Line 12-15 Revise language and grammar. Line 38-39 Specify how these suggested triggers are significant. What kinds of adverse events according to the GTT protocol, could these triggers contribute to identifying? **Page 18** Line 32-36 Revise language. Line 12-13 Please elaborate earlier in the paper, the problems the teams find related to the method being subjective. What significance did they mean that this has? Please elaborate earlier in the paper to what extent (all teams or just a few) did the teams experience, and how did they experience that they made changes in their performance of the GTT.

	Line 16 Please specify earlier in the paper on what they considered this task to be, and how they performed it. That will help to understand that they found it difficult, and if they had understood the task according to the protocols demands. Line 18 The logical foundation for the conclusion that the GTT method should be regarded as a qualitative tool should be strengthened, or the conclusion be revised.

REVIEWER	Christian von Plessen, MD PhD Senior physician Project Lead Safer Hospital Program Associate professor Department of Pulmonary and Infectious Diseases North Zealand Hospital DK-3400 Hillerød Denmark
REVIEW RETURNED	20-May-2013

GENERAL COMMENTS	The GTT is widely applied to measure harm to patients admitted to a hospital. Meanwhile the method has been criticized for low reproducibility and agreement of the harm rates that GTT review teams find. The study investigates reasons for these shortcomings. The researchers performed focus group interviews with GTT teams to explore their experiences with the review process. Between three and five team members participated in the interviews. Nurses and physicians were represented in all groups. The analyses of the interviews yielded the following themes: "Usefulness and use of the GTT", "Triggers", "Preventability of harm", "Team design", "Team tasks", "Team members' knowledge development" and "Documentation". The authors describe strengths and weaknesses of the GTT in light of each category and exemplify their conclusions with citations from the interviews. Important findings were that the review process changed over time, that the teams saw a need for nursing related triggers and that the

	teams found it difficult to give feedback to clinicians.
	
	The research question is relevant and the study original. Also the study design is suitable for the research question. Meanwhile the manuscript is not ready for publication. Mainly, the structure of the introduction is not clear and methods are not described in sufficient detail for the reader to understand. The citations from the interviews should be revised and/or expanded to give a better understanding of how the data support the categories and the conclusions. Moreover the discussion should address. Finally, language and syntax of the manuscript need to be improved to better convey the messages of the study. Thus, I suggest a major revision of the manuscript before it can be considered for publication.
	
	I have a couple of comments and examples from the manuscript to illustrate potential improvements:
	
	• Wording of title not clear: “Experiences of…?”
	• Introduction lacks focus and stringency.
	    o Difficult to understand from first paragraph of the introduction why study is important
	• Background of the researchers who analysed the interviews should be given to provide the reader with a better understanding of how categories “emerged”
	• The composition of the focus groups needs explanation (and discussion). Why were the focus groups conveyed as teams? Why not across teams?
	• Would suggest to summarise the questions for the focus group in a table or fact box
	• Would suggest to add the seven categories to results section in the abstract
	• It is not clear, why the results are presented dichotomised as ‘strengths’ and ‘weaknesses’
	• More detail needed from interviews to illustrate how categories emerged
	    o E.g.,, page 10, line 45: not really clear how citation illustrates that RN’s rather reviewed charts of the own and why others rather reviewed charts from

	other specialties • Page 11 line 23: What does “physician oriented” mean? • Points made in discussion are not always clear. o E.g., not clear what “…gained a greater understanding of how health care worked..” (page 18, para 2) means. o E.g., is it really a limitation of the study that teams had worked together for a long time? What about the fact that the study was conducted in Sweden where the GTT was modified to include preventability? What about generalizability of the findings to other settings? • Conclusions are not really plausible from the results presented. o For example, it is not clear why the GTT should be considered a qualitative tool while it originally was designed to monitor harm rates, ergo was introduced as a quantitative method. o It can be difficult to draw conclusions implying a ranking, such as “most important”, in qualitative studies. Does the data really support this conclusion? • The authors should use references in a critical and more selective manner. Are 42 citations really necessary? o E.g., reference number one (Classen et al) does not give an estimate of the prevalence of harm in Europe • Language and structure of the manuscript need revision to increase precision o For example, it is not clear how sentence two and three relate to sentence one in para 2 page 9. o Another example is on page 10, para 1: why mention the categories and what was considered most important here. Is the latter not better placed in the discussion section?

	○ What dose the conclusion "…useful, but subjective…" mean? ○ The abstract and the article summary need to be revised to give a better overview of the study. • Table 1: first part on level of harm not needed could be moved to text. Second part on preventability should remain, because it is a special feature of the Swedish adaptation of the GTT • Layout Table 2: Heading does not fit with columns

REVIEWER	Brian Bjørn Senior Consultant Danish Society of Patient Safety I declare that I have no competing interests.
REVIEW RETURNED	21-May-2013

GENERAL COMMENTS	The Institute for Healthcare Improvement Global Trigger Tool is to this date not well characterized, and the present study aims at expanding our knowledge of the tool. My recommendation for this manuscript is a major revision. The reasons for this recommendation are given below and are divided in two sections: A section of general comments and a section of more detailed comments. **General comments** My two main concerns about this manuscript are the lack of a clearly defined research question and a conclusion that is not sufficiently derived from the findings. The abstract states that the aim of the study was "to describe experiences of working with the Global Trigger Tool from team member perspectives." In the background section this is further expanded as "[t]his knowledge could provide us with a better understanding of the causes of disagreement between teams, and should provide valuable knowledge for refining the GTT." The authors need to clarify the research question(s) that they want to answer with this study. There is a huge difference between describing team member experiences and finding explanations for disagreement between teams. When the research question has been more clearly defined, this should guide the remainder of the manuscript: explain why these informants were chosen, who

	analyzed data and how, and what are the conclusions. The authors also need to more clearly state their own presuppositions and assumptions regarding the research question as this could influence the conduct of the interviews and interpretation of these. The conclusion in its present version is not supported by the data: 1. A qualitative focus-group interview is not the appropriate method for evaluating the measurement characteristics of the GTT. What can be learned from the present study is only what the team members think of the measurement characteristics. The appropriate method for characterizing the measurement would be studying inter-rater reliability and/or doing a test-retest study. The authors have previously published their findings from a reliability study, and the current conclusion probably spilt over from that study and from the authors presuppositions. 2. The study demonstrates (like the authors' reliability study) that these particular teams made gradual changes to the methodology; the authors then use this fact to conclude that GTT cannot be trusted as a quantitative measure. A more obvious conclusion would be this: If the method is not used correctly, it also does not work correctly. We can only conclude that the method does not work correctly if we have tested the actual method and not a derivative of it. **Specific comments** The following comments refer to the COREQ criteria for reporting qualitative research and to the RATS guidelines for reviewing ditto. **COREQ** The following numbers refer to the 32-item checklist. Items not mentioned here are considered sufficiently well reported in the manuscript. Item 1-5: The personal characteristics of the interviewer and the two moderators should be reported at least to the same level of details as for the informants (Table 2). Item 6-7: The relationship with informants should be reported. Item 10: Depending on the revised research question, the sampling strategy should be reported. Why were the GTT teams chosen as informants? Why not general clinical staff, department or hospital leadership, or regional quality staff? Why were teams kept together rather than mixed? Could mixing the teams have given rise to different discussions in the focus groups? Item 11: Method of approach should be reported Item 13: Did all members of the five teams take part or did anyone decline the invitation? Item 15: Presence of non-participants Item 17: Only the opening question is referred to in the manuscript.

	Were other questions used? If so, were they pilot tested prior to the actual study? Item 22: Data saturation – were informants recruited until no new knowledge was being obtained Item 23: Were informants given the option of commenting on transcripts? Item 25: Coding is insufficiently described. It is stated that co-authors discussed the categories – what form did that discussion have? Did any categories change after the discussion? Did the discussion make use of the transcripts or were they an exchange of views? Item 26: Were themes identified in advance or did they emerge? Item 28: Co-authors discussed the categories. Did informants get a chance to comment or correct? **RATS** RATS do not raise new questions, but the authors are encouraged to review the guidelines and keep these questions in mind when revising the manuscript. **Language** The manuscript would benefit from copyediting. There are currently too many sentences that are not clear and precise and a few that are incomprehensible. Examples: Institute for Health Care Improvement should be Institute for Healthcare Improvement. ”GTT implies that medical records […] are retrospectively reviewed […]. To imply something is to express it indirectly. The sentence therefore makes no sense. ”The GTT method is based on the study of retrospective structured medical records reviews […]” -> ”The GTT is a method for retrospective structured review of medical records […]” ”When reviewing records from their own area there was a risk for the RNs becoming jaded […]” It is not clear to me what this means? Did they ignore harm that occurred on their own wards? ”The teams found the triggers’ foreseeability intuitive, covering wide areas and facilitating reviewing.” What is meant by foreseeability of a trigger? ”Team design” should probably be team composition. ”The teams considered nurses more exacting in their documentation […]” I would think that what is meant is that nurses are more accurate and precise. Exacting does not carry that meaning. It

	describes something that makes severe demands. ”Epicrises” is not generally understood and should be replaced with discharge letter. The two words are also not synonyms. Adverse event vs. harm: Adverse event is used indiscriminately in the literature to mean both the event and its consequence. More precise terms are patient safety incident (for the event) and harm (for its consequence). The GTT uses the term harm instead of adverse event and the manuscript should follow that convention. **Reference 26** The study that is referenced is a newer Danish study. Unfortunately, this study uses a flawed statistical analysis that does not support its strong conclusions. The authors are advised to carefully evaluate the study.

VERSION 1 – AUTHOR RESPONSE

Referee 1:
The research question is defined as "describeexperiences of working with GTT". Reading the article one gets the impression that the objective is to learn from experienced GTT teams with the intent of improving the GTT method. Please clearify the aim.
We agree that the research question was vague and have clarified the aim. As a consequence we have changed the title accordingly.

The interviewer who did the focus group interviews should be described in more detail. Especially one needs to know more about the interviewers views on GTT before the study.
We have described the interviewers experience in the “Data collecting” section.

The manuscript would benefit from a review from a native English speaking person.
The revised manuscript has undergone language editing by a native English speaking proof reader.

The message is clear, that GTT should be considered a useful qualitative method, and that GTT is of less use as a quantitative method for measuring Changes in patient harm over time. I do not think that this
conclusion is supported by the study findings. The study design does not allow for strong conclusions on the quantitative properties of GTT.
We agree and have removed this conclusion.

Furthermore, the study finds that the review teams had modified the review methodology significantly over time. This invalidates any conclusions with regards to the "original" GTT methodology.
We believe that our teams are not alone in making modifications when using the GTT method. We point out different methodology used by the teams, and conclude this may have contributed to interrater
disagreement between teams. All teams maintained that they reviewed according to the GTT method as described in the manual, and that they only made adaptions to suit practical needs. We also point out that it takes time to become skilled and that as a reviewer one must learn to look at AEs from the

patient’s point of view. The results are not affected
by these changes in methodology, but are important and might contribute to inter-rater disagreement.
In the revised version we have tried to be more nuanced in our conclusions, and by doing so believe that they are valid also in regard to the “original” GTT methodology. We believe it
important to highlight the problems of reviewers making changes when
using the method in practice.

Referee 2
The context for how the teams have performed the GTT, is not made clear. At what level in their organizations have they performed the method and with what mandate? How have the teams been trained and followed up, to ensure that they perform according to the original protocol and definition? Such information is relevant for understanding how the teams experiences are relevant to other GTT teams.
We agree and have clarified this in
the “Participants” section as well as
in Table 2.

The Swedish GTT protocols differs from the original in that it considers the preventability of AE’s. This element has raised points of concern amongst the teams which do not apply to the original GTT protocol. The article should differentiate clearly which experiences that apply to the GTT protocol, and which experiences that apply to the modified version. Some of the teams experiences have been made within contexts that differ from the purpose of the original GTT protocol. The relevance of these experiences should be specified separately from those that have been made when sticking to the original protocol and purpose.
We have tried to separate the team’s opinion concerning preventability, that as the reviewer points out has been added to the Swedish GTT version. However, the teams' opinions on this may be valuable for readers from other countries. Some teams had experiences of reviewing records from hospital deaths. We have no reason to believe that experiences concerning the strengths and weaknesses of the GTT method from those Reviews are different from other reviews, as mortality might be included in
randomly chosen admissions. We, however, conclude that the GTT review methodology can be used in contexts other than randomly chosen admissions to hospital.

The study lacks questions related to the how the core measurements of the GTT protocol have been analyzed and presented. GTT is tailored for
analyzing the data with statistical process Control (SPC). What are the teams experiences with interpreting the data using SPC? The research questions could be presented more explicitly in the paper.
Thank you for this point of view. Our aim was to gain a deeper understanding of how the teams used the GTT in their Daily practice. None of the surveyed teams used SPC. None of the teams brought up the numerical results (level of AEs over time) of the reviews, and this aspect was not included in our questions. This is a limitation, and we have added this aspect to “limitations”.

There is a weak relation between the themes which the paper highlights as raised by the teams and the themes highlighted in the papers conclusion
We have tried to clarify the themes and have also revised the conclusions.

Page 2 Line 5
Please be more specific about the aims of describing the experiences.
We have tried to be more specific about the aims of describing the team’s experiences.

Line 25

The teams have evaluated their experiences with the GTT, and they have not evaluated the strengths and weaknesses of the GTT according to the protocols aims and standards. There is a difference here which the authors should be clear about. To evaluate the GTT according to its aims and standards would have required a more standardizing approach, than which
is applied in this study.
Thank you for pointing this out. The teams were asked in the introductory question about their experiences of the strengths and weaknesses of the GTT method. We have added this to the “Data collection” section.

Line 38-45
The first sentence does not specify which difference it refers to, although one can guess that it is the teams, it should be specified. The second sentence lacks logical foundation in the paper.
The first sentence; We have tried to clarify this.
The second sentence; We have removed the part concerning GTT as a qualitative tool.

Page 3 Line 30
This sentence lacks logical foundation in the paper and should be revised as suggested elsewhere in this review
We have removed the part concerning GTT as a qualitative tool.

Line 40-41
This sentence lacks foundation and reasoning in the paper.
We agree and have removed it.

Page 6 Line 23-25
Please clarify. In this sentence it seems like finding triggers is an aim, more than a mean, to find the AE’s
Thank you. We have revised the sentence.

Page 11 Lines 10- 21. It is not clear if the nurses registered the findings as AE’s according to the GTT, even if the findings did not comply with the AE definition, or if they registered the findings in other ways.
Please clarify.
The nurses did not determine the presence of AEs or not. They
handed over charts with positive triggers and their notes about potentially associated AEs to the physician for final judgment. We have tried to be clear about this in the description of the GTT method.

Page 13 Line 10-14
Please explain how the comment is relevant to the GTT protocol.
The preventability is added to the Swedish version of the GTT. The
teams were critical to this and we feel that this finding may be
important to others. We are also very clear about the fact that preventability has been added to the Swedish version, and not included in the original GTT method.

Page 14 Line 7-14
It seems if the tool has been used to facilitate dialogue related to the single events rather than to make measurements, based on statistical process control, as GTT is tailored for in the original protocol.
As described above, none of the surveyed teams used SPC. Instead, they went back to the clinics and discussed their findings. However, all hospitals calculated the level of AEs over time as part of the hospital patient safety program and this was done based on the team reviews. We have added

the missing aspect of the calculation of AEs over time in "Limitations".

Page 15: Line 45-47 needs clarification.
We have clarified this in the “Result” section.

Page 16
Line 14
Please specify for what and how, the teams found the GTT method useful.
We have tried to clarify this in the “Result” section, by reporting statements from the teams.

Line 30-40
It is not clear if the teams report that they have changed over time, how they supposedly perform the GTT according to the protocol, or if they have
chosen to use a deliberately modified version. A clarification of this is important for understanding the consistency of the method and how it is
conditioned.
We have tried to clarify this in the “Discussion” section.

Line 39
Referring to a team that obviously does not follow the protocol, may reflect differences in teams capability to apply the protocol. This should
probably also be discussed.
We have tried to explain that team’s point of view in the “Discussion”.

Line 43
Is a punctuation mark lacking?
We have added the punctuation mark.

Line 43 -48
It should be emphasized that the adjustments which Kennerly describes, were made before the method was implemented so that the review could be conducted consistently over time. This facilitated a good interrater reliability amongst reviewers (K = 0.62).
Thank you for your observance. We have removed the sentences about
Kennerly.

Line 54 -56
The reasoning behind the conclusion lacks some components.
1. Changes in methodology. The paper says Little about on what conditions the teams have changed their practice of the methodology.
2. Assessment of preventability. This element does not apply to the original GTT protocol. Conclusions on how the original GTT protocol
should be used can not be based on experiences with an element which it does not contain.
1. We have added examples (not reviewing in pairs and exceeding the time limit) of the teams' changes in methodology in the “Results” and “Discussion”.
2. We have removed the conclusion of GTT as only a qualitative tool. We have also added preventability as a possible limit in the
“Discussion”.

Page 17 Line 3-4 It is possible to suggest that the GTT method can produce information on AE’s that is useful in qualitative enquiry. But that does not exclude it’s original purpose, as the authors have concluded.
We agree and have removed statements concerning quantitative

value.

Line 12-15
Revise language and grammar.
The revision has undergone language editing by an native
English speaking proof reader.

Line 38-39
Specify how these suggested triggers are significant. What kinds of adverse events according to the GTT protocol, could these triggers contribute to identifying?
Thank you for this point. From the focus group interviews we Believe that elimination can indicate urinary retention as an easily forgotten injury. The trigger "pain" can indicate several AEs, i.e. postoperative bleeding.

Page 18 Line 32-36
Revise language.
The revised manuscript has undergone language editing by an native English speaking proof reader.

Line 12-13
Please elaborate earlier in the paper, the problems the teams find related to the method being subjective. What significance did they mean that
this has?
Please elaborate earlier in the paper to what extent
(all teams or just a few) did the teams experience,
and how did they experience that they made changes
in their performance of the GTT.
We have added a sentence about the Swedish judgment for
preventability in “Results” and “Discussion”. The only Connection between the perceived subjectivity is to the judgment of preventability, and we have tried to be very clear about this in our result and discussion sections. Concerning the teams’ changes in performance, we have described this in the beginning of the “Results” section. We have also clarified that changes
were made by all teams.

Line 16
Please specify earlier in the paper on what they considered this task to be, and how they performed it. That will help to understand that they found it
difficult, and if they had understood the task according to the protocols demands.
We have tried to clarify this in the “Results” and in the “Discussion”.

Line 18
The logical foundation for the conclusion that the GTT method should be regarded as a qualitative tool should be strengthened, or the conclusion be
revised.
We have removed the conclusion of GTT as a qualitative tool.

Referee 3
Mainly, the structure of the introduction is not clear and methods are not described in sufficient detail for the reader to understand.
In the revised version we have rewritten the introduction and hope that it has been clarified. We have also tried to describe the methods
in more detail.

The citations from the interviews should be revised and/or expanded to give a better understanding of how the data support the categories and the
conclusions.
We agree and have revised and expanded the citations.

Moreover the discussion should address.
Thank you, we have revised the “Discussion” part.

Finally, language and syntax of the manuscript need to be improved to better convey the messages of the study.
The revised manuscript has undergone language editing by an native English speaking proof reader.

Wording of title not clear: “Experiences of…?”
We have revised the title.

Introduction lacks focus and stringency. Difficult to understand from first paragraph of the introduction why study is important.
We have tried to be clearer about the importance of the study.

Background of the researchers who analysed the interviews should be given to provide the reader with a better understanding of how categories
“emerged”
We have added the experiences of the researchers who performed the
interviews.

The composition of the focus groups needs explanation (and discussion). Why were the focus groups conveyed as teams? Why not across teams?
We have tried to clarify this in the “Discussion” section.

Would suggest to summarise the questions for the focus group in a table or fact box.
Thank you for your suggestion. The point is important but as the other
referees did not point this out, we have chosen to write the questions
in the text.

Would suggest to add the seven categories to results section in the abstract.
We have added the categories to the result in the abstract.

It is not clear, why the results are presented dichotomised as ‘strengths’ and ‘weaknesses’
Thank you for this observation. The teams were asked in the introductory question about their experiences of strengths and weaknesses of the GTT method. We have added this explanation to the “Data collection” section.

More detail needed from interviews to illustrate how categories emerged
E.g.,, page 10, line 45: not really clear how citation illustrates that RN’s rather reviewed charts of the own and why others rather reviewed charts from other specialties.
We agree and have tried to clarify this in the “Analysis of the
interviews”.

Page 11 line 23: What does “physician oriented” mean?
We agree and have changed this to actions undertaken by physicians.

Points made in discussion are not always clear. E.g., not clear what “…gained a greater

understanding of how health care worked..” (page 18, para 2) means. E.g., is it really a limitation of the study that teams had worked together for a long time? What about the fact that the study was conducted in Sweden where the GTT was modified to include preventability? What about generalizability of the findings to other settings?
Thank you. We have added “..gained a greater understanding of
the healthcare structure”.
We agree and have moved the sentence of teams working together
for a long time to the “Discussion”.
We have tried to be clear about the conclusions when it comes to
preventability and conclusions that have no correlation to this Swedish
addition. We have also added the Swedish preventability judgment as
a possible limitation.
Thank you, but in qualitative research generalising is often not possible. One can only present statements from the surveyed teams and hopefully find things that other teams recognise and find useful.

Conclusions are not really plausible from the results
presented. For example, it is not clear why the GTT should be
considered a qualitative tool while it originally was
designed to monitor harm rates, ergo was introduced as a quantitative method. It can be difficult to draw conclusions implying a ranking, such as “most important”, in qualitative studies. Does the data really support this conclusion?
We agree and have removed the conclusion of GTT as a qualitative
tool.

The authors should use references in a critical and more selective manner. Are 42 citations really necessary? E.g., reference number one (Classen et al) does not give an estimate of the prevalence of harm in Europe.
We agree and have removed some of the references.

Language and structure of the manuscript need revision to increase precision. For example, it is not clear how sentence two and three relate to sentence one in para 2 page 9. Another example is on page 10, para 1: why mention the categories and what was considered most
important here. Is the latter not better placed in the discussion section?
What dose the conclusion “…useful, but subjective…” mean? The abstract and the article summary need to be revised to give a better overview of the study.
We have tried to clarify and become more precise.
We have removed "the most important task" in the beginning of the “Result” section. The reason for mentioning the categories was that
we wanted to provide a short summary of the categories found.
We have clarified that subjectivity relates to the Swedish judgment of
preventability.
We have revised the abstract and
the article.

Table 1: first part on level of harm not needed could be moved to text. Second part on preventability should remain, because it is a special feature of the Swedish adaptation of the GTT.
We have revised the Level of harm
in Table 1.

Layout Table 2: Heading does not fit with columns
We have changed the format of Table 2.

Referee 4
The authors need to clarify the research question(s) that they want to answer with this study. There is a huge difference between describing team member experiences and finding explanations for disagreement between teams. When the research question has been more clearly defined, this should guide the remainder of the manuscript: explain why
these informants were chosen, who analyzed data and how, and what are the conclusions. The authors also need to more clearly state their own
presuppositions and assumptions regarding the research question as this could influence the conduct of the interviews and interpretation of these.
Thank you. We have revised the research question. We have also tried to write in a manner that will guide the reader trough the article, e.g. why the informants were chosen and who analysed the data. We have written about the researchers' (who carried out the interviews) experiences of GTT and qualitative methods in “Data collection” .

The conclusion in its present version is not supported by the data: 1. A qualitative focus-Group interview is not the appropriate method for
evaluating the measurement characteristics of the GTT. What can be learned from the present study is only what the team members think of the measurement characteristics. The appropriate method for characterizing the measurement would be studying inter-rater reliability and/or doing a testretest study. The authors have previously published
their findings from a reliability study, and the current conclusion probably spilt over from that study and from the authors presuppositions.
2. The study demonstrates (like the authors’ reliability study) that these particular teams made gradual changes to the methodology; the authors then use this fact to conclude that GTT cannot be trusted as a quantitative measure. A more obvious conclusion would be this: If the method is not used correctly, it also does not work correctly. We can only conclude
that the method does not work correctly if we have tested the actual method and not a derivative of it.
1) We agree and have tried to clarify our aim with the study i.e. to describe strengths and weaknesses, from the perspective of team members working with the Global Trigger Tool method.
2) We agree and have removed the sentence “GTT should be seen as a
qualitative method”.

The following numbers refer to the 32-item checklist. Items not mentioned here are considered sufficiently well reported in the manuscript.
Thank you.

Item 1-5: The personal characteristics of the interviewer and the two moderators should be reported at least to the same level of details as for
the informants (Table 2).
We have written about the researchers' (who performed the interviews) experience of GTT and qualitative methods in the “Data collection” part.

Item 6-7: The relationship with informants should be reported.
We agree that dependency among researchers and informants may be
problematic and important to report, in particular when professionals perform research on patients. In this case, there is no vulnerable group and there is no dependency among the persons who conducted the interviews and informants. Therefore, we did not find it important to write about.

Item 10: Depending on the revised research question, the sampling strategy should be reported. Why were the GTT teams chosen as informants?
Why not general clinical staff, department or hospital leadership, or regional quality staff?

Why were teams kept together rather than mixed? Could
mixing the teams have given rise to different discussions in the focus groups?
We have tried to explain our sample strategy, i.e. our choice to
use existing teams both in the “Method” and the “Discussion”.

Item 11: Method of approach should be reported We have reported the study approach in the “Method” section.

Item 13: Did all members of the five teams take part or did anyone decline the invitation?
All team members took part in the interviews. We have added this
information to “Participants”.

Item 15: Presence of non-participants
There were no non-participants.

Item 17: Only the opening question is referred to in the manuscript. Were other questions used? If so, were they pilot tested prior to the actual study?
We have added examples of other questions in “Data collection”. The
questions were not pilot tested.

Item 22: Data saturation – were informants recruited until no new knowledge was being obtained?
As described by Krueger and Casey, three focus group interviews
are sufficient. We have added this in the “Analysis of the interviews”.

Item 23: Were informants given the option of commenting on transcripts?
The interviews were sent back to the participants for comments. We
have added this in the “Data collecting” part.

Item 25: Coding is insufficiently described. It is stated that co-authors discussed the categories – what form did that discussion have? Did any
categories change after the discussion? Did the discussion make use of the transcripts or were they an exchange of views?
The categories were changed until we reached consensus. We have tried to clarify the analytical process in the part called “Analysis of the interviews”.

Item 26: Were themes identified in advance or did they emerge?
The themes emerged during the analysis of the data. We have added
this in the “Analysis of the interviews”.

Item 28: Co-authors discussed the categories. Did informants get a chance to comment or correct?
Thank you for this point. The categories were not sent back to the
teams as by Krueger and Casey. This should only be done if it is of
importance, because participants are more limited in seeing the
larger picture.

The manuscript would benefit from copyediting. There are currently too many sentences that are not clear and precise and a few that are
incomprehensible. Examples:
Institute for Health Care Improvement should be Institute for Healthcare Improvement.
”GTT implies that medical records [...] are retrospectively reviewed [...]. To imply something is to express it indirectly. The sentence therefore
makes no sense.

”The GTT method is based on the study of
retrospective structured medical records Reviews [...]” -> ”The GTT is a method for retrospective structured review of medical records [...]”
”When reviewing records from their own area there was a risk for the RNs becoming jaded [...]” It is not clear to me what this means? Did they ignore harm that occurred on their own wards?
”The teams found the triggers’ foreseeability intuitive, covering wide areas and facilitating reviewing.” What is meant by foreseeability of a
trigger?
Thank you for your attention, wehave changed to Healthcare
Improvement.
We agree and have changedaccordingly.
We have removed “The GTT is based on…”
We have revised the sentence of nurses becoming jaded.
We have revised the sentence of “The teams found the triggers…”
”Team design” should probably be team composition.
We agree and have revised accordingly.

”The teams considered nurses more exacting in their
documentation [...]” I would think that what is meant is that nurses are more accurate and precise. Exacting does not carry that meaning. It describes something that makes severe demands.
We agree and have revised accordingly.

”Epicrises” is not generally understood and should be replaced with discharge letter. The two words are also not synonyms.
Thank you. We agree and have revised to discharge letter.

Adverse event vs. harm: Adverse event is used indiscriminately in the literature to mean both the event and its consequence. More precise terms are patient safety incident (for the event) and harm (for its consequence). The GTT uses the term harm instead of adverse event and the manuscript should follow that convention.
Thank you. In the GTT handbook from 2009, both AEs and harm are
mentioned. We chose the expression AE.

VERSION 2 – REVIEW

REVIEWER	Ellen Catharina Tveter Deilkås MD PhD Clinical Consultant/ Senior Scientist Akershus University Hospital Health Services Research I have no competing interests.
REVIEW RETURNED	30-Jul-2013

THE STUDY	1.The manuscript could elaborate more on possible differences in how the teams have been developed, trained, coordinated and followed up. Such descriptions could give reference to under what circumstances the results could be transferable to other hospitals. 2.The study has not brought to focus what experiences the teams have had with analysis and presentation of the GTT data. The GTT is tailored for analyzing the results in Statistical process control charts. That is why the method is based on small samples randomly and consecutively selected, with small time intervals. The discussion

	should elaborate more on why these experiences were not brought to focus in the interviews. Is it because the teams in Sweden do not analyze the GTT data in timeseries with statistical process control?
RESULTS & CONCLUSIONS	1. The title, text, conclusion and key messages should state more clearly that the results are not generalizable, and may be transferred to other settings, only under similar circumstances as those described in the study. For example; The title could write: 'Observed strengths and weaknesses of working with the Global Trigger tool….' rather than 'Strengths and weaknesses of working with the Global Trigger tool….' Likewise, the first sentence in the abstract page 2, lines 4-5 could write 'The aim was to describe strengths and weaknesses observed by team members working with the Global trigger Tool method' rather than 'The aim was to describe the strengths and weaknesses, from the perspective …' 2.The first category identified, 'Usefulness and use of the GTT', would probably be improved by splitting it in two. 'Usefulness' of the GTT relates to strengths and weaknesses for how GTT can be used, i.e. the method in itself. 'Use of GTT' could perhaps be renamed 'Application of the GTT', and relates to strengths and weaknesses related to how the GTT method has been applied; if it has been applied correctly, if the definitions have been correctly applied, how the triggers function, according to purpose, etc. 3.The teams are cited on that they miss opportunities in the GTT method to consider omission of care, especially nursing care. It was not made clear if the teams understood that accounting for omission of care would contradict the GTT definition of harm. It is important to know if the teams had a clear understanding of the GTT definition of harm. The article should clarify the authors interpretation of this element in the discussion. Page 10 lines 20-40 and page 12 line 30 – 45. 4.There is a contradiction between the teams claiming to comply with the GTT manual, while they apparently do not do so, considering for example the routine of both nurses reviewing the same charts independently. Which they do not do. Has development of such deviance not been addressed at annual meetings? The risk of developing different routines that deviate from the original manual is a weakness of doing the GTT procedure, but not a weakness of the procedure in itself. Page 10 line 50-55. 5.It is not made clear in describing the background, that the GTT method is primarily tailored for information use at hospital level, rather than department level. The teams believe that their most important task is to feed results back to their clinics, and say at the same time that relevance of the data is lacking since they reflect hospital level. This should be commented further in the discussion. A relevant question for the focus interviews would have been to what extent the GTT information is demanded by hospital leadership.
GENERAL COMMENTS	Page 4 line 34: The article referenced as number 17 is based on a study amongst oncological patients, which makes the study not representative for the GTT method. The GTT method randomly selects patients from the whole somatic population from a hospital (except pediatrics, and rehabilitation). The article shows that the GTT may not be valid for oncological patients, but the reviewing teams are not independent (page 1) since they have the same

	physician reviewer (page 2). I suggest that this reference is not included. Specific comments Page 4 line 12: rephrase Pge 5 lines 27-35: review language for more clarity Page 6 lines 22-23. Elaborate in the discussion on findings from this study, also related to interpretation of GTT definition. Page 17 line 29-34: Too long sentence. Reformulate. Page 19 line 25: Reformulate 'reach'. The title could write: 'Observed strengths and weaknesses of working with the Global Trigger tool….' rather than 'Strengths and weaknesses of working with the Global Trigger tool….' Likewise, the first sentence in the abstract page 2, lines 4-5 could write 'The aim was to describe strengths and weaknesses observed by team members working with the Global trigger Tool method' rather than 'The aim was to describe the strengths and weaknesses, from the perspective …' 1. The manuscript could elaborate more on possible differences in how the teams have been developed, trained, coordinated and followed up. Such descriptions could give reference to under what circumstances the results could be transferable to other hospitals. 2. The first category identified, 'Usefulness and use of the GTT', would probably be improved by splitting it in two. 'Usefulness' of the GTT relates to strengths and weaknesses for how GTT can be used, i.e. the method in itself. 'Use of GTT' could perhaps be renamed 'Application of the GTT', and relates to strengths and weaknesses related to how the GTT method has been applied; if it has been applied correctly, if the definitions have been correctly applied, how the triggers function, according to purpose, etc. 3. The teams are cited on that they miss opportunities in the GTT method to consider omission of care, especially nursing care. It was not made clear if the teams understood that accounting for omission of care would contradict the GTT definition of harm. It is important to know if the teams had a clear understanding of the GTT definition of harm. The article should clarify the authors interpretation of this element in the discussion. Page 10 lines 20-40 and page 12 line 30 – 45. 4. There is a contradiction between the teams claiming to apply to the GTT manual, while they apparently do not comply with it, considering for example the routine of both nurses reviewing the same charts independently. Has development of such deviance not been addressed at annual meetings? The risk of developing different routines that deviate from

	the original manual is a weakness of doing the GTT procedure, but not a weakness of the procedure in itself. Page 10 line 50-55. 5. It does not come clear as an understanding through the article, that the GTT method is primarily tailored for information use at hospital level, rather than department level. The teams believe that their most important task is to feed results back to their clinics, and say at the same time that relevance of the data is lacking since they reflect hospital level. This should be commented further in the discussion. A relevant question for the focus interviews would have been to what extent the GTT information is demanded by hospital leadership. 6. The study has not brought to focus what experiences the teams have had with analysis and presentation of the GTT data. The GTT is tailored for analyzing the results in Statistical process control charts. That is why the method is based on small samples randomly and consecutively selected, with small time intervals. The discussion should elaborate more on why these experiences were not brought to focus in the interviews. Is it because the teams in Sweden do not analyze the GTT data in timeseries with statistical process control? **Minor issues** Page 4 line 34: The article referenced as number 17 is based on a study amongst oncological patients, which makes the study not representative for the GTT method. The GTT method randomly selects patients from the whole somatic population from a hospital (except pediatrics, and rehabilitation). The article shows that the GTT may not be valid for oncological patients, but the reviewing teams are not independent (page 1) since they have the same physician reviewer (page 2). I suggest that this reference is not included. **Specific comments** Page 4 line 12: rephrase Pge 5 lines 27-35: review language for more clarity Page 6 lines 22-23. Elaborate in the discussion on findings from this study, also related to interpretation of GTT definition. Page 17 line 29-34: Too long sentence. Reformulate. Page 19 line 25: Reformulate ‘reach’.

VERSION 2 – AUTHOR RESPONSE

Reviewer's comments Authors response/solution.
1.The manuscript could elaborate more on possible differences in how the teams have been developed, trained, coordinated and followed up. Such descriptions could give reference to under what circumstances the results could be transferable to other hospitals.
In Table 2 we give some descriptions about the teams but we have also tried to elaborate more about this issue in the method part.
2.The study has not brought to focus what experiences the teams have had with analysis and presentation of the GTT data. The GTT is tailored for analyzing the results in Statistical process control charts. That is why the method is based on small samples randomly and consecutively selected, with small time intervals. The discussion should elaborate more on why these experiences were not brought to focus in the interviews. Is it because the teams in Sweden do not analyze the GTT data in timeseries with statistical process control?
We have added the fact that we did not discussed statistical process control as a limitation. We have tried to elaborate about this in the discussion.
1. The title, text, conclusion and key messages should state more clearly that the results are not generalizable, and may be transferred to other settings, only under similar circumstances as those described in the study.
We have added the following in the key message part and under limitations: "However, it was not clear whether their experiences of working with the GTT method reflected the opinions of teams from other hospitals or under other circumstances".
2.The first category identified, 'Usefulness and use of the GTT', would probably be improved by splitting it in two. 'Usefulness' of the GTT relates to strengths and weaknesses for how GTT can be used, i.e. the method in itself. 'Use of GTT' could perhaps be renamed 'Application of the GTT', and relates to strengths and weaknesses related to how the GTT method has been applied; if it has been applied correctly, if the definitions have been correctly applied, how the triggers function, according to purpose, etc.
We agree and have split the category "Usefulness and use of the GTT" into two categories.
We have also tried to elaborate about how the teams have applied the GTT method and about their perception of harm, in the discussion part.

3.The teams are cited on that they miss opportunities in the GTT method to consider omission of care, especially nursing care. It was not made clear if the teams understood that accounting for omission of care would contradict the GTT definition of harm. It is important to know if the teams had a clear understanding of the GTT definition of harm. The article should clarify the authors interpretation of this element in the discussion.
Page 10 lines 20-40 and page 12 line 30 – 45.
We agree and have added some text to the discussion part, concerning the team's opinion and understanding of the harm definition.
4.There is a contradiction between the teams claiming to comply with the GTT manual, while they apparently do not do so, considering for example the routine of both nurses reviewing the same charts independently. Which they do not do. Has development of such deviance not been addressed at annual meetings?
The risk of developing different routines that deviate from the original manual is a weakness of doing the GTT procedure, but not a weakness of the procedure in itself.
Page 10 line 50-55.
We have tried to clarify this in the method and the discussion parts.
We have also, by dividing the category "use and usefulness of the GTT" into two parts, tried to clarify the differences between the method in itself and the use of the method by the teams.

5.It is not made clear in describing the background, that the GTT method is primarily tailored for information use at hospital level, rather than department level. The teams believe that their most important task is to feed results back to their clinics, and say at the same time that relevance of the data is lacking since they reflect hospital level. This should be commented further in the discussion. A relevant question for the focus interviews would have been to what extent the GTT information is demanded by hospital leadership.
We have added that the GTT method primarily is intended for hospital use, both in the background and in the description of the GTT in the Method part.

We have discussed the fact that the team thought that their most important task was to bring back the results to the clinics, in the discussion part.

Unfortunately, we did not ask the question to the teams about to what extent the GTT information is demanded by hospital leadership.
Page 4 line 34: The article referenced as number 17 is based on a study amongst oncological patients, which makes the study not representative for the GTT method. The GTT method randomly selects patients from the whole somatic population from a hospital (except pediatrics, and rehabilitation). The article shows that the GTT may not be valid for oncological patients, but the reviewing teams are not independent (page 1) since they have the same physician reviewer (page 2). I suggest that this reference is not included.
Thank you for your suggestion but we have chosen to keep the reference number 17 as we find this study important as it analyses the GTT methodology in a manner that to the best of our knowledge is not earlier being done. Blant-Altman analysis is a well-known method for comparing different methods. However, we have added the fact that the study was made in an oncological context, as this is an important limitation.
Page 4 line 12: rephrase
We have rephrased this line.
Pge 5 lines 27-35: review language for more clarity
We have changed the example of triggers and hope this will be clearer.
Page 6 lines 22-23. Elaborate in the discussion on findings from this study, also related to interpretation of GTT definition. We have elaborated about this issue in the discussion part.
Page 17 line 29-34: Too long sentence. Reformulate.
We have divided the sentence into two.
Page 19 line 25: Reformulate ‘reach’.
We have changed from “reach” to “achieve a picture of”.