PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validity and reliability of a medical record review method identifying transitional patient safety incidents in merged primary and secondary care patients' records
AUTHORS	van Melle, Marije; Zwart, D.L; Poldervaart, Judith; Verkerk, Otto Jan; Langelaan, Maaike; van Stel, Henk; de Wit, Niek

VERSION 1 – REVIEW

REVIEWER	Carl de Wet Clinical Lead, Healthcare Improvement Unit, Queensland Health; Senior lecturer, School of Medicine, University of Griffith, Australia
REVIEW RETURNED	05-Feb-2018

GENERAL COMMENTS	5th February 2018 Dear Editor, **RE: Validity and reliability of clinician-based medical record review identifying transitional patient safety risks in merged primary and secondary care patients' records** Thank you for the opportunity to review this manuscript. There is considerable interest in developing psychometrically sound instruments to reliably assess the safety of health care. The authors developed and studied an adapted form of retrospective record review, which is widely considered to approximate a gold standard – at least when attempting to quantify the incidence of iatrogenic error or harm. As such, their experiences should be of particular interest to patient safety practitioners. However, I have four main reservations about the manuscript that you may wish to consider in reaching a decision about suitability for publication: 1. **Sample size**: No rationale for selecting the sample size was provided. The sample size seems small for a study that aims to determine the reliability of a method with some degree of precision. This perception is not only informed by the total number of records that were reviewed (n=301), but by the relatively short period of time and content that would have been reviewed in each record. In addition, only a small proportion (n=30, 10%) of records were reviewed twice. The primary outcome of this study (i.e. the reliability of the record review method) is only based on this sub set of reviews. 2. **Inter-rater reliability (IRR):** The main finding of the study, namely that IRR was poor, is comparable with the international literature. However, my understanding is that

each individual clinician performed a second review of *only 5 records*. If these records had instead been used to educate and align reviewers at the beginning of the study (as is customary in many studies of this type), it is possible that the IRR may have measurably improved. It would also have been useful to compare the findings and IRR of the two clinician groups (hospital specialist vs generalist).

3. **Validity**: One of the stated main aims of this study was to determine the validity of the adapted form of record review. It was unclear to me whether and how this aim was achieved. In particular, the discussion section does not include any reference to validity.

4. **Context, relevance and implications of the main study findings**: Reliable measurement is only one of the potential uses of the retrospective record review method. There is now also evidence of its usefulness in informing initiatives to improve the quality and safety standards of care and as an educational approach to raise awareness of safety-critical issues within health care systems and teams. Detecting TSIs, as in this study, therefore have potential benefits beyond simply quantifying clinical performance. This aspect could be more clearly articulated in the introduction and/or discussion. The authors may also wish to provide additional justification for recommending greater reliability of a method, given that this typically requires commensurate increases in resources. Assuming more TSIs could be detected in a feasible and reliable manner, the sheer number of TSIs may arguably exceed the ability of health care organisations to appropriately learn from and prevent them. Finally, there are already several (larger) studies that have examined the reliability of different adaptations of retrospective record review in different settings. In fact, Foster et al previously reported that 'acceptable' IRR would require that records be reviewed by three clinicians. While this study adds to the patient safety knowledge base, it does not meaningfully extend our understanding of the reliability (or validity) of this method. On the other hand, the findings about the number and type of TSIs are novel, interesting and potentially useful, but are not fully reported. These findings have important implications for clinical handover processes, yet they are not considered at all in the discussion section.

There are a number of minor grammatical errors throughout the manuscript. A few examples are provided below, with the caveat that they may appear pedantic or simply reflect my personal preferences.

- Title and manuscript – 'clinician-based'. Arguably, the review is not 'based' in a clinician, but performed by a clinician
- Title – 'risks' – throughout the manuscript, the authors use the term 'incident' (TSI) instead. This consistency should be reflected in the title
- Page 3, line 31 – 'records on': consider 'records for'
- Page 4, line 50: the sentence is incomplete
- Page 6, line 14 – patient's: patients' or patients
- Page 15, line 54 – been be: delete 'be'
- Page 15, line 57 – did not select on: rephrase
- Table 1b – the 'statements' are results, rather than

	demographic type data. It should be a separate table. Please let me know if I can be of any further assistance,

REVIEWER	Elaine Walsh University College Cork, Ireland
REVIEW RETURNED	16-Feb-2018

GENERAL COMMENTS	The comments that follow are also attached as a Word document. I recommend that the Word document be used as the formatting has not been retained here. Please note where a "NO" was checked in initial part of the review checklist above, the study may have met the requirement in part, but not in full. Explanations are provided below. This is an interesting study assessing a method for identifying transitional safety incidents (TSI) in digital records combining general practice and hospital information. As this method is being used to determine the primary outcome of an intervention study being conducted by the group, I feel that this paper would benefit from additional information being provided on the primary study to add context and promote clarity. Additionally, I feel that more detail should be provided regarding the specific aims and objectives of this study. It is somewhat unclear as to whether the aim of the study is to assess the use of the newly developed incident identification form by a panel of clinicians from different specialties to identify TSIs or whether the digital combined record is also being evaluated. It would be helpful to present more detail on the digital transitional records, particularly in relation to quality and missing information and how this may have impacted validity and reliability. Specific comments pertaining to individual sections of the paper are listed below: Abstract: The abstract differs from the structure outlined in instructions for authors which is as follows: o objectives: clear statement of main study aim and major hypothesis/research question o design: e.g. prospective, randomised, blinded, case control o setting: level of care e.g. primary, secondary; number of participating centres. Generalise; don't use the name of a specific centre, but give geographical location if important o participants: numbers entering and completing the study; sex and ethnic group if appropriate. Clear definitions of selection, entry and exclusion criteria o interventions: what, how, when and how long (this can be deleted if there were no interventions) o primary and secondary outcome measures: planned (i.e. in the protocol) and those finally measured (if different, explain why) – for quantitative studies only o results: main results with (for quantitative studies) 95% confidence intervals and, where appropriate, the exact level of statistical significance and the number need to treat/harm. Whenever possible, state absolute rather than relative risks o conclusions: primary conclusions and their implications, suggest

	areas for further research if appropriate. Do not go beyond the data in the article The study design is not described in the abstract, nor is information on participants. The abstract should mention the main study and state that the method of TSI review assessed in this study is being used to determine the primary outcome of the main intervention study. Strengths and limitations: Regarding bullet point 1: “In this study, we assessed validity and reliability of a transitional medical record review method in a transitional medical record, containing both the medical record of the general practitioner and hospital” This point describes what is being assessed in the study rather than describing a strength or limitation. Regarding bullet point 4: “As no accomplished reference standard is available in transitional patient safety, we chose a suboptimal reference standard, which could have led to an overestimation of our” Suggest omitting the word “accomplished”. The end of the sentence is missing. Regarding bullet point 5: “Validity and reliability was possibly affected by our heterogeneous reviewer group from both general practice and hospital, the limited number of the use of random pairs for the interrater reliability” The meaning of the latter part of the sentence referring to the use of random pairs is unclear. Introduction: I feel that more detail should be provided on the main study to provide greater clarity and context for this study. The introduction would benefit from a summary statement of the specific aims and objectives of this study. It is unclear currently if the aim of this study is solely to assess the actual method used to identify TSIs or if the study also aims to assess the performance of the novel digital transitional care record developed by the group. Regarding the sentence: “Patient safety is most often assessed through identification of safety incidents in medical record review study [10-12]” Perhaps reword as “Patient safety is most often assessed through identification of safety incidents in review of medical records” Methods: Information regarding the study design should be provided at the beginning of the methods section. Is this study a retrospective review? What reporting protocol does the study adhere to? Patient selection:

	• Can the exclusion criteria be explained in more detail? How was a "complete" transitional record defined? Missing information from the transitional records is subsequently noted as a finding in the results/discussion sections. • Were the patients included selected from the intervention or control groups of the main study? • What method of randomisation was used? • The actual numbers of patients who were included/excluded may be better placed in the results rather than the methods section. • Figure 1 contains formatting errors and requires a legend as it is not clear what the abbreviations mean. Linking medical records form primary and hospital care: Who is the trusted third party? Was this the same for the creation of all transitional records or did it vary? If conducted by different people did the quality of the records vary? Was an assessment of standardisation and record quality conducted by the research team? Reviewers: Were the reviewers involved in the care of the patients whose records they were reviewing? You mention the reviewers attending a reflection meeting during the data collection period. Was this the data collection period for the main study or does it refer to the time period allocated to the reviewers to identify the TSIs? When were the reviewers allocated the records and what was the time frame for identifying the TSIs? The review process: • "The review process was based on methodology from previous medical record review studies [14,15] which shaped the development of a systematic identification form for TSIs, and the process of testing and refining, as well as the reviewers' training". Regarding "testing and refining": What did this involve? Was the identification form piloted? • Were clinicians responsible for patient care informed regarding clinically significant TSIs that were detected by the reviewers? • What was the randomisation process used to allocate the records to the 6 reviewers? Reliability: How were the records that had double review allocated to the 6 reviewers? You mention having 15 pairs of reviewers in the discussion section. Did some reviewers review more records than others? Was inter-rater reliability calculated for each pair of reviewers-could you provide additional information on how this was conducted? Validity: As the construct validity was not being assessed in this study, this perhaps should be omitted from the methods section. Regarding concurrent validity: the reference standard was 3 particular TSIs identified in the records by the research team. Could

	you include the rationale for choosing these particular TSIs and describe composition and expertise of the research team? Results: Application of the Kruskal Wallis and Pearson Chi square tests is mentioned in the methods section but the actual test results are not specified in the results section. Could these test results be included? Median age is described in the text but the mean is presented in Table 1a In Table 1b there is text missing from the medical specialty section. In Table 1b in the section pertaining to "Statement Results Reviewers", a mix of numbers and text are presented as results. Could this be clarified? Regarding missing information and overall record quality: was there any assessment of the records by the research team? As the reviewers were required to extract data as the first step in the process, could availability of information and quality of the record have impacted on the results of the review process? Regarding reliability: did the authors consider having all reviewers review the same sample of records and then conducting inter-rater reliability? Regarding: "Despite of the instructions to apply the definition of TSI strictly and conservatively, they inclined to subscribe the small failings as minor imperfections instead of TSIs or as consequences of the incomplete research database" How was this established? Can numbers be given for this? It is mentioned in the methods that a relationship between reviewer characteristics and identification of TSIs was sought by using the Pearson Chi Square but this is not presented in the results section. Discussion: Main findings: Regarding the statement: "In the current form, clinician-based transitional medical record review proved not to be a reliable method for TSI identification. Although TSIs were identified in one in seven care transitions and 92% of these proved correctly labelled as TSI, reviewers missed 78% of objectively identifiable TSIs". This statement refers to validity not reliability. Regarding the statement: "The inter-rater reliability was low and reviewers significantly differed in frequency and characteristics of the identified TSIs". There is no mention of differing TSI characteristics in the results section. Could the quality of the records and the requirement for data extraction account for low inter-rater reliability? Comparison to literature:

	It should be clarified what type of medical records were being reviewed in the cited studies and what instruments were being used for detection of transitional safety incidents for the purposes of comparison. Regarding the study including “all patients visiting hospital and general practice”. It should be stated here if any of the patients in this study had received the intervention of the primary study. Regarding: “However, because of these different backgrounds, they had different, often contrasting views on TSIs and the healthcare system”. This is not mentioned in the results section. How was this ascertained? If clinicians are not best placed to establish significance could the authors make some suggestions as to who should conduct this task-trained researchers, allied healthcare professionals?? The composition of the research group identifying the TSIs in the study would be of relevance here. Strengths and Limitations: “Concerning limitations all double medical records were randomly distributed among the other reviewers, leading to 15 different pairs of reviewers (instead of specific reviewer-pairs), which might have decreased inter-rater reliability” Having 15 pairs of reviewers was not mentioned in the results section. “Finally, correspondence of discharge and visits to the outpatient clinic was regularly missing” How many of the records were missing this information and how was this determined? Could this information be included in the results section? Implications: Selecting a more homogenous group of reviewers might improve reliability but not necessarily validity and hence may not provide a solution.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1) *"Sample size: No rationale for selecting the sample size was provided. The sample size seems small for a study that aims to determine the reliability of a method with some degree of precision. This perception is not only informed by the total number of records that were reviewed (n=301), but by the relatively short period of time and content that would have been reviewed in each record. In addition, only a small proportion (n=30, 10%) of records were reviewed twice. The primary outcome of this study (i.e. the reliability of the record review method) is only based on this sub set of reviews."*

We understand the concern of the reviewer and we probably were not clear enough on the purpose of the current study. This study was part of a pilot to test the research method for a large intervention study, improving transitional care. For this main study, we did perform a sample size calculation. However, for this current study we had to make concessions as our time and budget were limited and

specialists are expensive. Therefore, the sample size of 300 medical records were the maximum number we could reach.

2) *Inter-rater reliability (IRR): The main finding of the study, namely that IRR was poor, is comparable with the international literature. However, my understanding is that each individual clinician performed a second review of only 5 records. If these records had instead been used to educate and align reviewers at the beginning of the study (as is customary in many studies of this type), it is possible that the IRR may have measurably improved. It would also have been useful to compare the findings and IRR of the two clinician groups (hospital specialist vs generalist).*

We agree with the reviewer on the potential of educating the reviewers; we actually had the intention to have a reflective meeting every week and discuss all the issues they encountered and differences in the reviews. However, practicality caught up with us and we could only organise one reflective meeting and one end meeting which all of the reviewers could attend within the time period. This was another problem with using specialists for review of a large amount of medical record. Among other reasons, this made us change the reviewers in our intervention study to final master year' medical students. With them, we were able to have weekly reflective meetings and discuss the differences in the doubly reviewed medical records. We did compare the hospital specialists versus generalists, and did find a difference between the two groups in the amount and type of TSIs identified (the GP group identified more and other TSIs than the hospital specialists). However because the results seemed very much relying on two outliers and the number of reviewers was only six, we did not feel comfortable publishing that in the article.

3) *Validity: One of the stated main aims of this study was to determine the validity of the adapted form of record review. It was unclear to me whether and how this aim was achieved. In particular, the discussion section does not include any reference to validity.*

We thank the reviewer for the helpful comment and changed the first paragraph of the discussion to clarify (p17 ‘Main findings’).

4) *Context, relevance and implications of the main study findings: Reliable measurement is only one of the potential uses of the retrospective record review method. There is now also evidence of its usefulness in informing initiatives to improve the quality and safety standards of care and as an educational approach to raise awareness of safety-critical issues within health care systems and teams. Detecting TSIs, as in this study, therefore have potential benefits beyond simply quantifying clinical performance. This aspect could be more clearly articulated in the introduction and/or discussion. The authors may also wish to provide additional justification for recommending greater reliability of a method, given that this typically requires commensurate increases in resources. Assuming more TSIs could be detected in a feasible and reliable manner, the sheer number of TSIs may arguably exceed the ability of health care organisations to appropriately learn from and prevent them. Finally, there are already several (larger) studies that have examined the reliability of different adaptations of retrospective record review in different settings. In fact, Foster et al previously reported that ‘acceptable’ IRR would require that records be reviewed by three clinicians. While this study adds to the patient safety knowledge base, it does not meaningfully extend our understanding of the reliability (or validity) of this method. On the other hand, the findings about the number and type of TSIs are novel, interesting and potentially useful, but are not fully reported. These findings have important implications for clinical handover processes, yet they are not considered at all in the discussion section.*

We thank the reviewer for the recognition of the importance and usefulness of this record study and helpful suggestion to accentuate this aspect in the introduction and discussion section. We adjusted the discussion section accordingly and underlined its clinical relevance (p17 ‘Main findings’ and p 20 ‘Implications’).

5) *Textual comments:*
There are a number of minor grammatical errors throughout the manuscript. A few examples are provided below, with the caveat that they may appear pedantic or simply reflect my personal preferences.
• Title and manuscript – 'clinician-based'. Arguably, the review is not 'based' in a clinician but performed by a clinician
• Title – 'risks' – throughout the manuscript, the authors use the term 'incident' (TSI) instead. This consistency should be reflected in the title
• Page 3, line 31 – 'records on': consider 'records for'
• Page 4, line 50: the sentence is incomplete
• Page 15, line 54 – been be: delete 'be'
• Page 15, line 57 – did not select on: rephrase

We adjusted the textual comments accordingly.

Reviewer 2

1) *This is an interesting study assessing a method for identifying transitional safety incidents (TSI) in digital records combining general practice and hospital information. As this method is being used to determine the primary outcome of an intervention study being conducted by the group, I feel that this paper would benefit from additional information being provided on the primary study to add context and promote clarity.*
Additionally, I feel that more detail should be provided regarding the specific aims and objectives of this study. It is somewhat unclear as to whether the aim of the study is to assess the use of the newly developed incident identification form by a panel of clinicians from different specialties to identify TSIs or whether the digital combined record is also being evaluated.
I feel that more detail should be provided on the main study to provide greater clarity and context for this study. The introduction would benefit from a summary statement of the specific aims and objectives of this study. It is unclear currently if the aim of this study is solely to assess the actual method used to identify TSIs or if the study also aims to assess the performance of the novel digital transitional care record developed by the group.

   We thank the reviewer for the kind words on our study and the suggestion to clarify how this study relates to the larger intervention study. This study is one of two pilot studies performed to test and improve the primary outcome measuring method of the main study. The first one titled 'Pilot study on identification of incidents in healthcare transitions and concordance between medical records and patient interview data' was previously published in the BMJ Open. According to the suggestion of the reviewer, we clarified this and added context to the current study (p6 last paragraph). We also specified our aims (p 7 last paragraph).

2) *It would be helpful to present more detail on the digital transitional records, particularly in relation to quality and missing information and how this may have impacted validity and reliability.*

   According to the reviewer's helpful comment, we added detail on the digital transitional records (p8) and its missing information (p14). Additionally, we have added this factor to the discussion setting (p17).

3) *The study design is not described in the abstract, nor is information on participants.*
The abstract should mention the main study and state that the method of TSI review assessed in this study is being used to determine the primary outcome of the main intervention study.

We added the requested information to the abstract as suggested.

4) *Regarding bullet point 1:*
"In this study, we assessed validity and reliability of a transitional medical record review method in a transitional medical record, containing both the medical record of the general practitioner and hospital" This point describes what is being assessed in the study rather than describing a strength or limitation.
Regarding bullet point 4:
"As no accomplished reference standard is available in transitional patient safety, we chose a suboptimal reference standard, which could have led to an overestimation of our" Suggest omitting the word "accomplished". The end of the sentence is missing.
Regarding bullet point 5:
"Validity and reliability was possibly affected by our heterogeneous reviewer group from both general practice and hospital, the limited number of the use of random pairs for the interrater reliability" The meaning of the latter part of the sentence referring to the use of random pairs is unclear.

We thank the reviewer for pointing out these obscurities and adjusted the bullet points accordingly

5) *Regarding the sentence: "Patient safety is most often assessed through identification of safety incidents in medical record review study [10-12]" Perhaps reword as "Patient safety is most often assessed through identification of safety incidents in review of medical records"*

We adjusted the textual comment accordingly (p6)

6) *Information regarding the study design should be provided at the beginning of the methods section. Is this study a retrospective review? What reporting protocol does the study adhere to?*

We thank the reviewer for this helpful comment and added information on the study design (p8 'Study design and patient selection'). We did try to find a reporting protocol to adhere to, but unfortunately, we could not find one for the validation of this type of measurement tool. The QUADAS-2 comes closest; it is designed for measurement tools, but only covers questionnaires, rendering it not of use.

7) *Patient selection:*
- *Can the exclusion criteria be explained in more detail? How was a "complete" transitional record defined? Missing information from the transitional records is subsequently noted as a finding in the results/discussion sections.*
- *Were the patients included selected from the intervention or control groups of the main study?*
- *What method of randomisation was used?*

We thank the reviewer for this helpful comment and added information to the patient selection paragraph (p8 last paragraph). As this study concerns a pilot before the actual intervention study, the patients were not part of either the intervention nor the control group.

8) *The actual numbers of patients who were included/excluded may be better placed in the results rather than the methods section.*

We placed the numbers of in- and excluded patients in the result section accordingly (p13 first paragraph).

9) *Figure 1 contains formatting errors and requires a legend as it is not clear what the abbreviations mean.*

We thank the reviewer for pointing that out, we made a large adjustment to the figure to clarify.

10) *Linking medical records form primary and hospital care: Who is the trusted third party? Was this the same for the creation of all transitional records or did it vary? If conducted by different people did the quality of the records vary? Was an assessment of standardisation and record quality conducted by the research team?*
The review process: "The review process was based on methodology from previous medical record review studies [14,15] which shaped the development of a systematic identification form for TSIs, and the process of testing and refining, as well as the reviewers' training".
Regarding "testing and refining": What did this involve? Was the identification form piloted?

To answer the reviewer's comment, we added information on the linking process accordingly (p9 'Linking medical records from primary and hospital care'). This study was prepared with several previous steps: first, we tested the technical linkage process to assess if it worked properly. Then, our research team together with the expert on medical record reviews (ML) tested the completeness of the medical records and usability of the measurement tool. We made adjustments accordingly and the majority of missing data was retrieved. Unfortunately, still some correspondence remained missing although it was not sure whether this was because of lack of registration by the healthcare professional/ organisation or due to the data extraction done for the research databases. The current manuscript describes the third and last phase: the testing of the medical record review method identifying TSI's. We chose only to report on the medical record review method because the three-phased pilot would be too much information. Regarding the record quality, this was assessed by our reviewers and we have added the information to the result section.

11) *Were clinicians responsible for patient care informed regarding clinically significant TSIs that were detected by the reviewers?*

As the patient medical records were anonymised, we could not inform the clinicians responsible for patient care on individual patients. We did inform them of our (group) findings and the type of TSIs that were found.

12) *What was the randomisation process used to allocate the records to the 6 reviewers?*

We added this information to the methods section accordingly (p10 'Reliability').

13) *Reviewers: Were the reviewers involved in the care of the patients whose records they were reviewing?*
You mention the reviewers attending a reflection meeting during the data collection period. Was this the data collection period for the main study or does it refer to the time period allocated to the reviewers to identify the TSIs? When were the reviewers allocated the records and what was the time frame for identifying the TSIs?

We thank the reviewer for pointing out these obscurities. To address the comments: the reviewers were independent specialists or generalists and did not have a treatment relationship with the included patients. We have added this information together with a specification on the reflection meeting to the manuscript (p9 'Reviewers'). The allocation and time frame are specified in the online appendix.

14) *Reliability: How were the records that had double review allocated to the 6 reviewers? You mention having 15 pairs of reviewers in the discussion section. Did some reviewers review more records than others? Was inter-rater reliability calculated for each pair of reviewers-could you provide additional information on how this was conducted?*

We appreciate the reviewer's questions and added more information on the allocation of the doubly reviewed medical records (p10 'Reliability')

15) *Validity: As the construct validity was not being assessed in this study, this perhaps should be omitted from the methods section. Regarding concurrent validity: the reference standard was 3 particular TSIs identified in the records by the research team. Could you include the rationale for choosing these particular TSIs and describe composition and expertise of the research team?*

We thank the reviewer for this suggestion and placed the part on the construct validity and previous published pilot in the introduction section (p7 last paragraph). We also added information on the purposeful choice of the three TSIs used as a reference standard (p11 last paragraph).

16) *Results: Application of the Kruskal Wallis and Pearson Chi square tests is mentioned in the methods section but the actual test results are not specified in the results section. Could these test results be included?*

As the results of all of these analyses were not significant, we only summarised the results in the result section and did not mention all the values as there were many characteristics (p13 'Patient and reviewer characteristics'). We believe adding them would provide too much information and therefore obscurity. Yet, of course we could add these data is the reviewers and editor would prefer as such.

17) *Median age is described in the text but the mean is presented in Table 1a. In Table 1b there is text missing from the medical specialty section.*

We adjusted table 1 accordingly

18) *Regarding missing information and overall record quality: was there any assessment of the records by the research team? As the reviewers were required to extract data as the first step in the process, could availability of information and quality of the record have impacted on the results of the review process? Could the quality of the records and the requirement for data extraction account for low inter-rater reliability?*

Our research assistants did first check the completeness of the medical records and we did have strict rules about what to identify as a TSI. Also, the quality of the records was similar for all the reviewers, so variation the quality of the records between reviewers should not have affected the inter-rater reliability. Having said that, we did see differences in applying the strict rules in records of lesser quality. We tried to get the reviewers in line there during the reflection meeting and in individual conversations. However, we do believe this variation in handling of records of lesser quality may have affected inter-rater reliability.

19) *Regarding reliability: did the authors consider having all reviewers review the same sample of records and then conducting inter-rater reliability?*

Indeed, we did consider as such. However, because of time constraints and lack of resources we did not execute as such. However, all reviewers reviewing the same sample of records might have improved our IRR.

20) *Regarding: "Despite of the instructions to apply the definition of TSI strictly and conservatively, they inclined to subscribe the small failings as minor imperfections instead of TSIs or as consequences of the incomplete research database" How was this established? Can numbers be given for this?*

We specified the sentence in the discussion section (p17 'Main findings'). The statement was based on a discussions during the reflection and discussion meetings, therefore, we cannot give any numbers for that.

21) *It is mentioned in the methods that a relationship between reviewer characteristics and identification of TSIs was sought by using the Pearson Chi Square but this is not presented in the results section.*
Regarding the statement: "The inter-rater reliability was low and reviewers significantly differed in frequency and characteristics of the identified TSIs". There is no mention of differing TSI characteristics in the results section.

We thank the reviewer for this helpful comment. We at first did perform this analysis and it showed differences between the GPs and hospital specialists and also between the younger and older reviewers. However, these results were mainly based on two outliers. And because of the small number of reviewers we decided not to present this analysis in the current article. Therefore, we deleted this sentence from the statistical analysis section. Additionally, we deleted the reference to this calculation from the discussion section.

22) *Main findings: Regarding the statement: "In the current form, clinician-based transitional medical record review proved not to be a reliable method for TSI identification. Although TSIs were identified in one in seven care transitions and 92% of these proved correctly labelled as TSI, reviewers missed 78% of objectively identifiable TSIs". This statement refers to validity not reliability.*

We specified the two separate statements accordingly (p17 'Main findings)

23) *Comparison to literature: It should be clarified what type of medical records were being reviewed in the cited studies and what instruments were being used for detection of transitional safety incidents for the purposes of comparison.*

We thank the reviewer for this suggestion and added this information to the comparison to literature paragraph accordingly (p17).

24) *Regarding the study including "all patients visiting hospital and general practice". It should be stated here if any of the patients in this study had received the intervention of the primary study.*

As described earlier, we clarified the place of the current study in the larger TIPP intervention study in the introduction section. As this study is a pilot before the start of the intervention study, the patients did not receive the intervention.

25) *Regarding: "However, because of these different backgrounds, they had different, often contrasting views on TSIs and the healthcare system". This is not mentioned in the results section. How was this ascertained?*

We thank the reviewer for the helpful comment and clarified this statement accordingly (p18).

26) *If clinicians are not best placed to establish significance could the authors make some suggestions as to who should conduct this task-trained researchers, allied healthcare professionals?? The composition of the research group identifying the TSIs in the study would be of relevance here.*

We thank the reviewer for the useful comment and added information to the discussion section (p20 first paragraph)

27) *Strengths and Limitations: "Concerning limitations all double medical records were randomly distributed among the other reviewers, leading to 15 different pairs of reviewers (instead of specific reviewer-pairs), which might have decreased inter-rater reliability" Having 15 pairs of reviewers was not mentioned in the results section.*

We understand the vagueness of our sentence and adjusted the sentence accordingly (p19).

28) *"Finally, correspondence of discharge and visits to the outpatient clinic was regularly missing". How many of the records were missing this information and how was this determined? Could this information be included in the results section?*

We thank the reviewer for this useful comment and added data on missing data to the results section (p14 last paragraph). As the missings can either be caused by data extraction problems or registration omissions on the healthcare professional's side, we do believe some of the missing data are indeed unsafe situations and therefore minor TSIs potentially leading to adverse events.

29) *Implications: Selecting a more homogenous group of reviewers might improve reliability but not necessarily validity and hence may not provide a solution.*

We partly agree with the reviewer's comment. Therefore, we adjusted the statement slightly as we believe it may provide part of the solution (p20).

We hope we have answered the reviewers' comments sufficiently.
Thank you for considering our work.

VERSION 2 – REVIEW

REVIEWER	Dr Elaine Walsh University College Cork, Ireland
REVIEW RETURNED	20-Apr-2018

GENERAL COMMENTS	**Bullet points:** Regarding bullet point 4: *"As no accomplished reference standard is available in transitional patient safety, we chose a suboptimal reference standard"* The meaning of the word "accomplished" and also the overall meaning of the statement is not clear. **Methods:** <u>Study design and patient selection:</u>

	The text pertaining to the definition of a complete medical record added by the authors in the revised submission is unclear. Point 2 appears to be a further explanation of Point 1 rather than an additional criterion. Perhaps this information would be more appropriately presented as a list of specific inclusion and exclusion criteria? The method of randomisation used for patient selection is not mentioned. <u>The review process:</u> Regarding the following point raised in the original review: • *"The review process was based on methodology from previous medical record review studies [14,15] which shaped the development of a systematic identification form for TSIs, and the process of testing and refining, as well as the reviewers' training".* Regarding "testing and refining": What did this involve? Was the identification form piloted? It is understandable that in the interest of being concise the authors are reluctant to present too much information on the process of developing their record review method. However, as testing this new method of review is the focus of this study, more information should be provided on the actual method. The information could be referenced if detailed in the authors' previous publication or included as an Appendix. <u>Validity:</u> Regarding the following point raised in the original review: • Regarding concurrent validity: the reference standard was 3 particular TSIs identified in the records by the research team. Could you include the rationale for choosing these particular TSIs? As this is the method for determining one of the 2 main study outcomes some further detail is desirable. Who determined that these TSIs were easily identifiable compared to other TSIs? If it was e.g. the medical record expert or a clinician/group of clinicians then it should be stated. Were these TSIs identified to the reviewers in their training? To take for example "Redundant diagnostic testing": it might not necessarily occur to a reviewer to label a blood test repeated within 3 months a TSI unless this was specifically highlighted to the reviewer in advance. **Results:** <u>Characteristics of identified transitional incidents</u> It is important to be clear about completeness of the included medical records as the authors report that 60% of the TSIs detected concerned "Information from hospital to general practice". The authors mention in their response that correspondence was missing

	from included records, that the reason for missing correspondence could not be determined in some cases and that missing correspondence could have constituted a TSI. It needs to be clearly stated whether or not the TSIs reported in the study could have been due to missing information arising from the process of combining the primary and secondary care records rather than being an actual TSI with the potential to impact patient safety. If there is uncertainty then it needs to be clearly stated for the reader as it significantly affects interpretation of the result, with the actual number of TSIs detected in the study being potentially much lower than the figure reported. <u>Validity:</u> Findings from "Individual and group discussions" are mentioned here. If qualitative findings are being reported in the results section, there should be a corresponding section in the methods outlining how data collection and analysis were conducted.

REVIEWER	Carl de Wet School of Medicine, University of Griffith, Gold Coast, Australia
REVIEW RETURNED	24-Apr-2018

GENERAL COMMENTS	Thank you for submitting a revised draft for further review. The clarifications and edits have improved the overall quality of the manuscript. You may wish to consider deleting the word 'patient' from 'patient records' to avoid duplication, but it is a matter of personal taste. Best wishes with the next stages of TIPP and with your work to improve the safety of clinical handover

VERSION 2 – AUTHOR RESPONSE

Please find enclosed the resubmission of our paper entitled: *"Validity and reliability of a medical record review method identifying transitional patient safety incidents in merged primary and secondary care patients' records.*"

In this new version we processed the very helpful comments of the reviewer. The text was clarified according to the provided suggestions, and information was added. Below you will find detailed answers to the subsequent comments per reviewer. Additionally, we adjusted the textual suggestions. In the manuscript we marked the altered text using track changes.

1) Regarding bullet point 4:

"As no accomplished reference standard is available in transitional patient safety, we chose a suboptimal reference standard"

The meaning of the word "accomplished" and also the overall meaning of the statement is not clear.

We thank the reviewer for pointing this out and adjusted the bullet point (p4).

2) Methods:

Study design and patient selection:

The text pertaining to the definition of a complete medical record added by the authors in the revised submission is unclear. Point 2 appears to be a further explanation of Point 1 rather than an additional criterion. Perhaps this information would be more appropriately presented as a list of specific inclusion and exclusion criteria?

We thank the reviewer for this suggestion. We restructured the text on the patient selection according to the reviewer's comment and think it is much clearer now (p8).

3) The method of randomisation used for patient selection is not mentioned.

We added information on the randomisation process accordingly (p8).

4) The review process: Regarding the following point raised in the original review:

• "The review process was based on methodology from previous medical record review studies [14,15] which shaped the development of a systematic identification form for TSIs, and the process of testing and refining, as well as the reviewers' training".

Regarding "testing and refining": What did this involve? Was the identification form piloted?

It is understandable that in the interest of being concise the authors are reluctant to present too much information on the process of developing their record review method. However, as testing this new method of review is the focus of this study, more information should be provided on the actual method. The information could be referenced if detailed in the authors' previous publication or included as an Appendix.

To provide more detail on the development of the medical record review method, we added a subsection named 'Review development' to the manuscript (p10)

5) Validity: *Regarding the following point raised in the original review:*

• Regarding concurrent validity: the reference standard was 3 particular TSIs identified in the records by the research team. Could you include the rationale for choosing these particular TSIs?

As this is the method for determining one of the 2 main study outcomes some further detail is desirable. Who determined that these TSIs were easily identifiable compared to other

TSIs? If it was e.g. the medical record expert or a clinician/group of clinicians then it should be stated. Were these TSIs identified to the reviewers in their training? To take for example "Redundant diagnostic testing": it might not necessarily occur to a reviewer to label a blood test repeated within 3 months a TSI unless this was specifically highlighted to the reviewer in advance.

We added extra information on the choice of TSIs in the methods (validity p12) accordingly.

6) Results: Characteristics of identified transitional incidents

It is important to be clear about completeness of the included medical records as the authors report that 60% of the TSIs detected concerned "Information from hospital to general practice". The authors mention in their response that correspondence was missing from included records, that the reason for missing correspondence could not be determined in some cases and that missing correspondence could have constituted a TSI. It needs to be clearly stated whether or not the TSIs reported in the study could have been due to missing information arising from the process of combining the primary and secondary care records rather than being an actual TSI with the potential to impact patient safety. If there is uncertainty then it needs to be clearly stated for the reader as it significantly affects interpretation of the result, with the actual number of TSIs detected in the study being potentially much lower than the figure reported.

We agree that the incompleteness of medical records could lead to an overestimation of identified TSIs. To clarify, we added information to the results (Validity section p15) and the discussion (strengths and limitation section p20-21). In the results, we added information on the sort of correspondence-related TSIs identified by the reviewers. Not all correspondence-related TSIs regarded missing correspondence, the type of TSI also included delayed, incomplete or incorrect correspondence, or correspondence not processed at the general practice. In the discussion section we elaborated further on the consequences of possible missing data for our research.

7) Validity:

Findings from "Individual and group discussions" are mentioned here. If qualitative findings are being reported in the results section, there should be a corresponding section in the methods outlining how data collection and analysis were conducted.

We did mention these discussions in the methods section, but only shortly. For clarity on the information acquisition, we added information to the methods section (p12)

We hope we have answered the reviewer's comments sufficiently.
Thank you for considering our work.