

BMJ Open Validity and reliability of a medical record review method identifying transitional patient safety incidents in merged primary and secondary care patients' records

Marije A van Melle,^{1,2} Dorien L M Zwart,¹ Judith M Poldervaart,¹ Otto Jan Verkerk,³ Maaike Langelaan,⁴ Henk F van Stel,¹ Niek J de Wit¹

To cite: van Melle MA, Zwart DLM, Poldervaart JM, *et al.* Validity and reliability of a medical record review method identifying transitional patient safety incidents in merged primary and secondary care patients' records. *BMJ Open* 2018;**8**:e018576. doi:10.1136/bmjopen-2017-018576

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-018576>).

Received 7 July 2017
Revised 20 June 2018
Accepted 20 July 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands

²Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

³Faculty of Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

⁴Netherlands Institute for Health Services Research (NIVEL), Utrecht, The Netherlands

Correspondence to

Dr Marije A van Melle;
mav41@medschl.cam.ac.uk

ABSTRACT

Objective Inadequate information transfer during transitions in healthcare is a major patient safety issue. Aim of this study was to pilot a review of medical records to identify transitional safety incidents (TSIs) for use in a large intervention study and assess its reliability and validity.

Design A retrospective medical record review study.

Settings and participants Combined primary and secondary care medical records of 301 patients who had visited their general practitioner and the University Medical Center Utrecht, the Netherlands, in 2013 were randomly selected. Six trained reviewers assessed these medical records for presence of TSIs.

Outcomes To assess inter-rater reliability, 10% of medical records were independently reviewed twice. To assess validity, the identified TSIs were compared with a reference standard of three objectively identifiable TSIs.

Results The reviewers identified TSIs in 52 (17.3%) of all transitional medical records. Variation between reviewers was high (range: 3–28 per 50 medical records). Positive agreement for finding a TSI between reviewers was 0%, negative agreement 80% and the Cohen's kappa –0.15. The reviewers identified 43 (22%) of 194 objectively identifiable TSIs.

Conclusion The reliability of our measurement tool for identifying TSIs in transitional medical record performed by clinicians was low. Although the TSIs that were identified by clinicians were valid, they missed 80% of them. Restructuring the record review procedure is necessary.

INTRODUCTION

Multidisciplinary care provided by professionals from different settings (eg, between general practice and hospital) is prone to patient safety incidents.^{1 2} Transitions between healthcare settings as well as concurrent care by medical specialists and general practitioners (GPs) can lead to 'transitional safety incidents' (TSIs). TSIs may range from unsafe situations like discrepancies between the patients' medical record in hospital or

Strengths and limitations of this study

- The transitional medical records database included all transitions from general practice to hospital and vice versa, allowing us to follow the patient's journey and measure patient safety crossing healthcare levels.
- We aimed to identify all potentially harmful incidents that indicate weaknesses in the system, not only adverse events.
- The reference standard used in this study is suboptimal. This is caused by the fact that there is no valid reference standard available for this purpose.
- Reliability was possibly affected by our heterogeneous reviewer group from both general practice and hospital, the limited number of reflection meetings, and the use of random instead of set reviewer pairs for the inter-rater reliability.

general practices to adverse events causing harm to the patient, such as unplanned readmissions or even death.^{2 3}

TSIs are common: after hospital discharge almost 50% of patients are confronted with a potential health-threatening situation, such as inadequate handover or medication errors.³ It is estimated that more than 25% of all readmissions may be prevented by improving the communication to the GP and the patient, and by improving the discharge process.⁴ Furthermore, concerning the referral process, specialists deem 10% of referrals as possibly inappropriate and only 16% of all referral letters are perceived as 'good'.^{5 6} As both referral and discharge letters contain information that is key to the patient's care, their 'loss in transition' by inadequate communication or inadequate information exchange processes leads to potentially unsafe care.^{7 8}

To contribute to the improvement of transitional patient safety, we designed the Transitional Incident Prevention Programme (TIPP). In short, in a controlled before–after study, we aim to assess the effectiveness of the multifaceted TIPP intervention targeting transitional patient safety between general practice and hospital, executed in both an urban and rural setting.⁹ As one of the primary outcomes for this intervention study we wanted to measure ‘transitional patient safety’ before and after the intervention.

Patient safety is most often assessed through identification of safety incidents in review of medical records.^{10–12} As other patient safety measurement methods, like incident reporting by healthcare professionals or questionnaires assessing patient experiences, are inadequate to count the actual number of incidents,¹³ we chose medical record review as the primary instrument to assess the effect of the TIPP intervention.

In a previous pilot study, we assessed construct validity by comparing our medical record review method with patient interviews and found a concordance of 64%.^{14 15} The current paper presents a second pilot study to further assess the measurement properties of the primary outcome in this intervention programme. In this study, we assessed the reliability and validity of identification of TSIs by record review in a patient record database with both primary and secondary care patient’s data, performed by independent primary and secondary care clinicians.

METHODS

Study design and patient selection

A retrospective medical record review study was performed, identifying TSIs in a research database linking patients’ primary and secondary care medical records. Although we also tested the database linking process prior to this pilot study, we only present the data on reliability and validity of the review method in the current study.

For this transitional medical record database, we included patients from the departments of cardiology and gastroenterology at the University Medical Center Utrecht, the Netherlands. Inclusion criteria were: (1) patients with at least one contact with one of the research hospital departments (admission or a visit to the outpatient clinic) between 1 April 2013 and 1 October 2013 and (2) patients registered with a referring GP affiliated with the primary care research database of the Julius Primary Care Network (an academic routine care registration network, with 300 000 patients from 200 GPs) in the city of Utrecht with at least one registration in the GP medical record within the study period. Patients were excluded if (1) the contact with the hospital department was only a consultation for another hospital department or (2) the patients’ medical records were missing a major part of one of the medical records, such as medical history, medication or consultations.

Of the eligible medical records, we selected a random sample of 301 patients, using computer-generated random numbers. As this study concerns a pilot study, we did not do a sample size calculation. We chose a number of 300 medical records based on practical issues such as resources and availability of reviewers.

Linking medical records from primary and hospital care

Currently, in the Netherlands hospitals and GPs work in separate electronic medical record systems. For this study, we constructed a digital transitional medical record database consisting of data from both general practice and hospital. According to current privacy regulations, a Trusted Third Party (Custodix) linked all patients’ records from both sources after pseudonymising the data.¹⁶ The anonymous transitional medical records contained: patients’ age and gender, admissions to the cardiology and/or gastroenterology departments, nursing documentation, and the following information from both the GP and the hospital: medical history, current medication, documented consultations, diagnostic examinations and correspondence. A digital timeline based on the collected medical records was created for each patient to summarise the patient journey. Unfortunately, the referral letters could not be obtained, as they were sent using an independent mailing system and therefore not stored in both medical records.

Reviewers

We chose a multidisciplinary group of six reviewers who reviewed the medical records individually. To cover all possible views on TSIs, we chose reviewers from all professions involved: three GPs, two cardiologists and one gastroenterologist. The reviewers were not attached to the research departments and did not have a medical relationship with the included patients. The reviewers received a half-day training beforehand on the medical record review process, the definition of TSIs and the context of the TIPP study. During the training, the reviewers practised with case histories and the transitional medical record database. During the data collection period, the reviewers could discuss problems encountered with the research team. As all reviewers reviewed random patients visiting different research departments (not only of their own specialty), they had the possibility to consult each other. They also attended one reflection meeting on the record review process with the whole group.

Review development

The review process was based on methodology from previous medical record review studies^{17 18} which shaped the development of a systematic identification form for TSIs, and the process of testing and refining, as well as the reviewers’ training. We took the review method and manual used in the in-hospital medical record review studies performed by Baines *et al* and Zegers *et al* as a basis.^{17 18} We used an iterative process to develop it into the review for this current study. First, all items of the

in-hospital review were reviewed and assessed on their fit for transitional patient safety. Items that did not fit transitional patient safety or could not be reformulated to fit transitional patient safety culture were deleted. Second, we reformulated questions to fit transitional patient safety. Third, we added items based on our knowledge from our previously published pilot.¹⁴ The resulting record review was then discussed in the research team and, after adjusting, discussed with a wider expert team with GPs, hospital specialists and patients. Finally, the record review was tested two times by MAVM, JMP, ML and DLMZ on actual patient medical records. The final version is presented in online supplementary appendices A and B.

Review process

Fifty transitional medical records were randomly assigned to each of our reviewers and instructions were given to identify TSIs during the study period using a computer-generated random sample using Microsoft Excel random number generation. When a TSI was identified, the reviewers summarised and classified the TSI according to type, cause, severity and preventability. The definition of TSI used was 'any unintended or unexpected event in patient care between different healthcare organisations which could have led or did lead to harm for one or more patients receiving care'.¹⁴ We used the Eindhoven Classification Model to characterise the cause¹⁹ and the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) to classify the severity of the TSI.²⁰ Preventability was classified using a 6-point scale from '(nearly) no evidence of preventability' to '(definitely) evidence of preventability'.¹⁷ The reviewers also assessed the quality of the medical records. For details of the review steps and classifications, see online supplementary appendices A and B.

Reliability

To determine inter-rater reliability, a random sample of 10% of the medical records was reviewed by two reviewers independently. To select these medical records, we used a simple computer-based random sample method. All reviewers rereviewed five medical records that were already reviewed by other reviewers. We did not assign set reviewer pairs, but each reviewer rereviewed medical records first reviewed by each of the other reviewers. To assess inter-rater variation, we collected reviewer characteristics: age, gender, medical specialty, working experience, scientific experience, experience in incident analysis and their opinion on three statements regarding incidents and transitional safety using a 5-point Likert scale from strongly disagree to strongly agree (table 1).

Validity

For validating our record review method, we focused on content and concurrent validity.

Content validity is defined as: 'the extents to which items appear 'on the face of it' to be measuring what the

instrument is intended to measure'.¹⁵ We first searched the literature and existing medical record review studies within organisations.^{17 18 21} Then, we used these existing intraorganisational record reviews to develop our review method identifying TSIs. Hospital specialists, GPs and an expert in medical record review studies (ML) were consulted to assess our record review method and we tested the usability. Subsequently, we asked the reviewers to judge our TSI identification method on completeness and clarity, both individually during the review process and in a group discussion at the end of the review process. We had frequent informal meetings with each of our reviewers during the review process and asked specifically for obscurities in the review method. These comments were collected in an excel sheet. During the final group discussion notes were taken.

Concurrent validity is defined as 'the extent to which scores on a new measure are related to scores from a criterion measure administered at the same time'.¹⁵ As no criterion measure exists within transitional patient safety,¹² we used a standardised assessment of three objectively identifiable TSIs as a reference standard, namely (1) presence and timeliness of correspondence from hospital to GP, (2) redundant diagnostic testing (ie, a diagnostic test unnecessarily performed by both GP and hospital in a short time period) and (3) communication of in-hospital prescription changes to GP. We purposefully chose these three types of TSI for their easy and systematically approachable identification process in the available medical records. These TSIs were chosen by the research team and medical record review expert because all three types proved well-documented processes in the medical record in the first pilot of the medical record review study.¹⁴ The three TSIs were also part of our six reviewers' training for the current study: the reviewer case report form (CRF) clearly stated these TSIs within the list of transitions and risks for TSIs at the start of each form and examples were given during the training for these TSIs, including how to find them in the medical record (online supplementary appendices B). The easily identifiable TSIs were systematically assessed by our research assistants (final year medical students) using a strictly protocolled step-by-step guide, as described in more detail in online supplementary appendix A. Additionally, to assess whether all reviewer-identified TSIs met the definition of a TSI, the research team assessed all reviewer-identified TSIs.

Statistical analysis

We calculated differences between the patients whose medical records were reviewed (n=301) and the patients of all eligible medical records (n=1399) using Kruskal-Wallis test for continuous variables and the Pearson χ^2 test for categorical variables. With descriptive statistics, we presented the total of TSIs and the number per reviewer, type, cause, severity and preventability. To assess the inter-rater reliability, we tested positive and negative agreements of identified TSIs and calculated Cohen's

Table 1 Patients' and reviewers' characteristics**(A) Patient characteristics (n=301)**

Age, median (range)	60 (19–93)
Male sex, N (%)	167 (56)
Medical specialty, N (%)	
General practice	301 (100)
Cardiology	211 (70)
Gastroenterology	90 (30)

(B) Reviewer characteristic (n=6)

Age, mean (range)	55 (32–72)
Male sex, N	3
Medical specialty, N	
General practice	3
Cardiology	2
Gastroenterology	1
Years of working experience as a medical specialist or GP, range	0.5–32
Carried out a scientific research before, N yes	5
Carried out a medical record research before, N yes	6
Experience with analysing medical incidents, N yes	1
Worked as a teaching physician, N yes	3
Statements presented to our reviewers (n=6)	
1: How useful is medical research focused on incidents, rated on a scale from 1 to 10, with 10 being the most positive, range	7–9
2: As healthcare is complex care transitions are not ever flawless. Yet, not all of these flaws represent safety incidents, range	Neutral to strongly agree
3: Incidents always harm the patient, range	Strongly disagree to disagree

GP, general practitioner.

kappa and intraclass correlation coefficient (ICC) model 2 in which Cohen's kappa disregards our varying reviewer pairs (observer nested within subjects) and the ICC disregards the binomial nature of our outcome.²² IBM SPSS Statistics V.21.0 was used for all statistical analyses.

All reviewers signed a confidentiality agreement.

Patient and public involvement

Patient advocates were involved in the design of the larger TIPP study; however, they were not directly involved in the current study. Because of the anonymous nature of

the medical record, individual cases were not communicated to the research department. However, the departments were informed of the study results.

RESULTS

Patient and reviewer characteristics

In total, 2069 patients were admitted or visited the outpatient clinics of the research departments (figure 1). Of those, 670 patients were excluded because they were not registered at an affiliated GP. Of the remaining eligible 1399 patients, 91 were excluded because of an incomplete medical record. Of the remaining 1308, we took a random sample of 301 patients. The 301 randomly selected patients were not significantly different from the total group of 1399 patients in the database. Their median age was 60 years (19–93) in the included population and 59 (11–93) in the total group of 1399 ($p=0.49$), and 56% vs 53% was male ($p=0.57$). Of the included patients, 70% visited the cardiology department. Table 1 presents the characteristics of the included patients' characteristics. Patient characteristics were evenly distributed between the reviewers, except for age. By chance, the age of the patients reviewed by reviewer III was significantly higher

Table 2 Inter-rater reliability in 30 transitional medical records reviewed by two reviewers independently

	Second reviewer		Total
	TSI identified	No TSI identified	
First reviewer			
TSI identified	0	7	7
No TSI identified	3	20	23
Total	3	27	30

TSI, transitional safety incident.

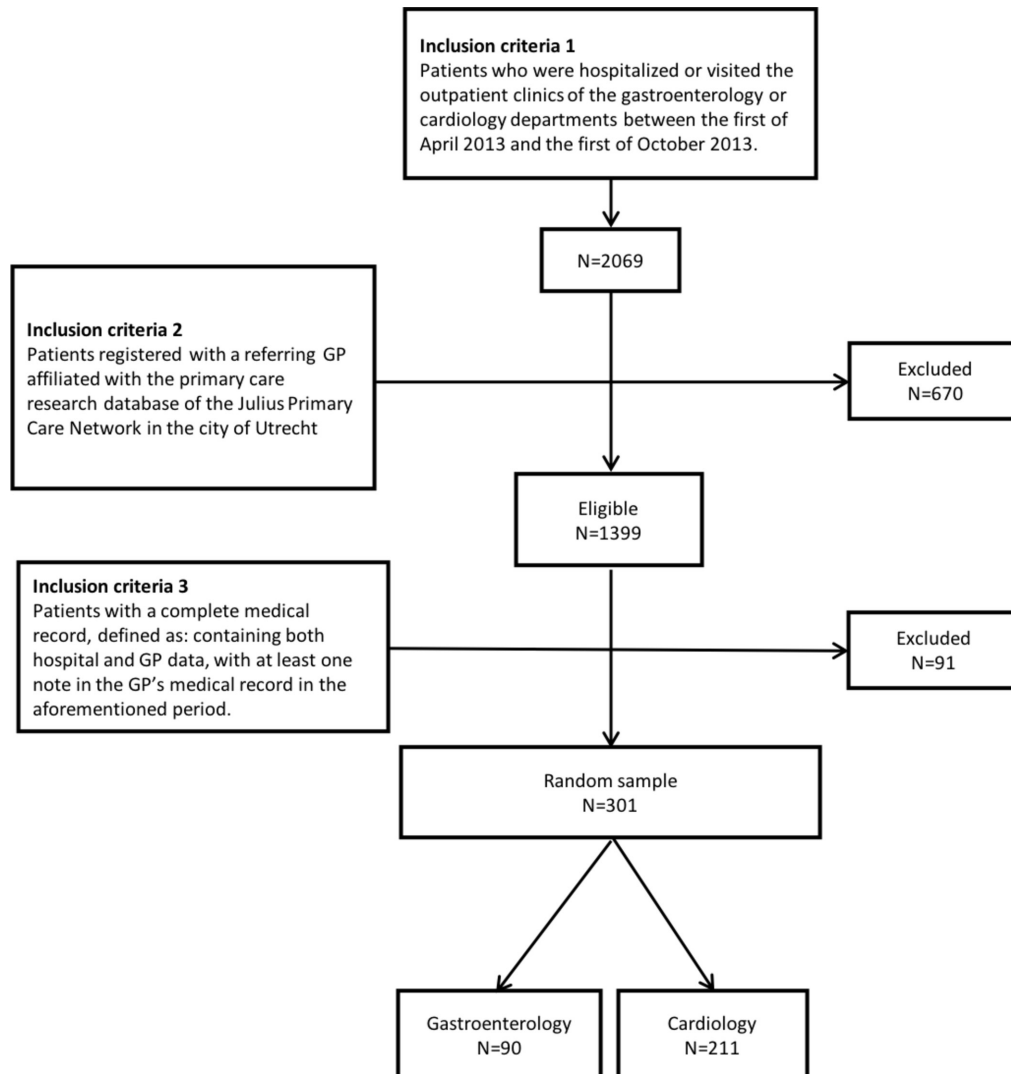


Figure 1 Flow chart of patient selection for the clinician-based medical record review study of the TIPP study. GP, general practitioner; TIPP, Transitional Incident Prevention Programme.

(median of 67 years). The age of the reviewers ranged between 32 and 72 and they had 0.5–32 years of working experience as a GP or specialist (table 1).

Characteristics of identified transitional incidents

In total, the 301 reviewed patient records revealed 422 transitions between general practice and hospital or vice versa (1.4 transitions per patient in 6 months). In 52 of 301 (17.3%) records, 60 TSIs were identified, which means 1 TSI in every 7 care transitions.

The main types of identified TSIs concerned ‘Information from hospital to general practice’ (n=35; 60% of total TSI), ‘Medication and other prescriptions’ (n=19; 32%) and ‘Registration in the medical record’ (n=16; 27%). The majority of TSIs were assigned to organisational (n=44; 73%) or human causes (n=20; 33%). No TSIs were judged to be patient related. As multiple types or causes could be attributed to a single TSI, these numbers add up to more than 100%. Of the TSIs concerning ‘Information from hospital to general practice’, 14 regarded missing outpatient correspondence, 8 regarded missing

discharge letters, 4 letters were delayed, 4 letters were incomplete, 3 incorrect and in 2 TSIs a letter was sent, but the information was not processed. Our reviewers assessed 48 (89%) TSIs as an unsafe situation (NCC MERP category A), 3 (5%) TSIs as a near miss that did not reach the patient (NCC MERP category B) and 2 (3%) TSIs as reaching the patient, of which 1 led to temporary harm (NCC MERP categories C and F). Severity of seven TSIs could not be assessed because of insufficient information. Thirty (55%) of TIs were assessed as probably or definitely preventable. Preventability of five TSIs could not be assessed because of insufficient information. The quality of the GP medical records was rated as ‘poor’ in 0.0%–10.2% of all reviewed medical records and as ‘good’ in 2.1%–43.8%. The quality of the hospital medical records was rated by the different reviewers as ‘poor’ in 0.0%–6.4% and as ‘good’ in 10.0%–60.4%. The quality of the correspondence between the hospital and the GP and vice versa was rated as ‘poor’ in 3.2%–42.9%, as ‘good’ in 0.0%–25.8% and as ‘not possible to review due to missing

Box 1 Examples of identified TSIs in the medical record review

- ▶ The hospital specialist stops two prescriptions, however, 2 months later the general practitioner (GP) repeats the old medication prescriptions.
- ▶ Medication in hospital's record and GP's record differs; (acenocoumarol vs phenprocoumon).
- ▶ In a discharge letter, two different cholesterol inhibitors are prescribed. The patient finds out and contacts his GP.
- ▶ No note of atrial fibrillation with moderate heart function and complaints of dyspnoea diagnosed in hospital in the medical record of the GP.
- ▶ The cardiologist's medical record does not mention diabetes diagnosed by the GP (other risk factors are registered).
- ▶ No note or letter of a hospital admission in the GP's medical record.
- ▶ The GP calls the hospital for the results of a colonoscopy in hospital because of persistent complaints. The GP did not receive correspondence of the results or treatment and the patient does not know.
- ▶ After discharge, the discharge letter asks the GP to check the patient's potassium level, however, the blood test is not performed.
- ▶ A patient has three types of potential potassium-elevating medication separately prescribed by the GP or the hospital physician, but the potassium blood level is never checked.
- ▶ The GP's medical record notes indicate doubt about the first responsible physician for treating the patient's high blood pressure (cardiologist or GP) in a patient with poor medication adherence. Because the patient prefers the GP, he takes up the treatment. Patient is discharged from the outpatient visits to the cardiologist.

data' in 0.0%–67.3%. In 84 medical records (28%), correspondence from hospital to the GP was missing, although signs of a letter being sent was found in either the GPs' or hospitals' records. However, we could not determine the cause (missing because of faulty data extraction or omission of registration at the healthcare practitioners' side). **Box 1** shows some examples of identified TSIs.

The frequency of TSIs identified per reviewer ranged from 3 to 28 in 75 transitions. Five of the six reviewers identified 3–10 incidents, and one identified a total of 28 TSIs. Patients' age or gender did not differ between those patients who did and those who did not experience a TSI. The patients who experienced a TSI were more frequently treated by the cardiologist (82.7%) compared with patients who did not experience a TSI (67.5%).

Reliability

In the 30 transitional medical records reviewed by two reviewers independently, 10 TSIs (33%) were found in total (table 2). The positive agreement was 0%, the negative agreement was 80%, Cohen's kappa was -0.15 and ICC was -0.18 .

Validity

Content validation: Reviewers reported neither missing items nor obscurity in the TSI identification method used. Yet, reviewers showed having a problem applying the definition of TSI to frequently occurring, 'small' transitional failings. Despite of the instructions to apply the definition

of TSI strictly and conservatively, individual and group discussions showed that they inclined to subscribe the small failings as minor imperfections instead of TSIs or as consequences of the incomplete research database. The fact that all but one reviewer agreed with the statement 'As healthcare is complex care, transitions are not ever flawless. Yet, not all of these flaws represent safety incidents' also indicated that reviewers had doubts about the content validity (table 1). The only reviewer who did not agree with the statement identified most TSIs.

Concurrent validation: Ninety-two per cent of the reviewer-identified TSIs met the researchers' definition of a TSI. Using the reference standard for TSIs research assistants identified 194 objectively identifiable TSIs in 179 (62%) medical records: of these, 153 regarded correspondence, 5 regarded redundant diagnostic testing and 36 reflected communication of prescription changes. Only 43 (22%) of these TSIs were identified by the reviewers and hence 151 (78%) TSIs were missed. Within reviewers, the percentage of objectively identifiable TSIs found ranged from 3% to 47%.

DISCUSSION

Main findings

In the current form, our transitional medical record review method proved not to be a reliable method for TSI identification. The inter-rater reliability was low and reviewers significantly differed in frequency of the identified TSIs. Face validity of the medical record review study seemed acceptable, as most TSIs that were identified in one in seven care transitions proved correctly labelled as TSI. However, concurrent validity proved to be low; reviewers missed 78% of objectively identifiable TSIs. Hence, the identification and categorisation of TSIs were undoubtedly subjective. The limited agreement with our reference standard and the high variability between reviewers points at a lack of a shared understanding of the definition of a TSI, despite of the format of the review procedure or the training. This might have been provoked by the reviewers' insecurity when handling medical records of lower quality.

The identified TSIs mainly concerned lacking or delayed correspondence and medication-related incidents, caused by an organisational or human factor. The majority of TSIs concerned unsafe situations, yet 1 of 60 (1.7%) identified TSIs caused patient harm.

Comparison to literature

The reliability of our review method was much poorer than in other medical record review studies. While the inter-rater reliability in medical record reviews by clinicians ranged between fair and substantial agreement (Cohen's kappa 0.24 – 0.71),^{12 23–25} we found a poor Cohen's kappa of -0.15 . These studies used various measurement tools, such as the Global trigger tool, the Harvard medical practice study method and a structured patient chart review for adverse drug events, to identify

mainly intraorganisational incidents within hospital medical records. Variation in patient selection, types of patient records and/or reviewers' characteristics may explain these differences. For instance, in a comparable study on identification of TSIs after discharge, Forster *et al* found a suboptimal inter-rater reliability (Cohen's kappa 0.26) in 319 case reports of patients with poor outcomes following discharge. Because these patients are known to have a high risk for TSIs, incidents would have been easier to identify and therefore the agreement probably was better than in our study in which we did not select high-risk patients only, but included all patients visiting hospital and general practice.²⁴ Second, in contrast to Forster *et al* in which hospital-based reviewers assessed only hospital medical records, our reviewers assessed medical records that originated from different organisations. Additionally, others may have worked with reviewers from the same discipline while we consciously chose to work with a heterogeneous group of reviewers in terms of medical disciplines and hoped this would add to the validity of the record review. However, we hypothesise that because of these different backgrounds, they had different, often contrasting views on TSIs and the healthcare system which may have resulted in the identification of different TSIs. The reference and end discussion sessions substantiated this belief. For, as much as we would like one healthcare system, the current healthcare system is fragmented in which hospitals and general practices have their own language and view on patient safety.²⁶ Together, it is likely that the variation in reviewer characteristics has hampered the reliability of our TSI identification process.²⁴

Our reviewers may have had problems identifying TSIs because they viewed incidents from a clinician's perspective instead of a patient safety perspective. Such general underestimation of incidents and harm by clinicians is described previously.²⁷ Clinicians deal with risks on a daily basis. They learn to be resilient and deal with unsafe situations instantly at their occurrence, thus making risk management part of routine practice.¹³ However, such attitude could lead to a loss of awareness about the state of the systems and processes that impact patient care, thus hampering identification of safety incidents.

Strengths and limitations

Our study has several strengths. First, the transitional medical records database included all transitions from general practice to hospital and vice versa. This adds to the content validity as we were able to follow the patient's journey and measure patient safety crossing healthcare levels.²⁸ As TSIs occur between the healthcare organisation, they are often not registered in the interorganisational medical record. In our previous pilot study, we concluded that the availability of the medical records at both ends of the patient journey was needed to identify TSIs looking for double information, discrepancies and textual clues.¹⁴ Second, many studies that measure patient safety and incidents only include adverse events;

we typically included all potentially harmful incidents that indicate weaknesses in the system.²⁹

Concerning limitations, although recommended by Forster *et al* to use three reviewers per medical record simultaneously, practical and financial limitations limited the number of reviewers, and therefore, each medical record was assessed by a single reviewer.²⁴ Furthermore, all double medical records were randomly distributed among the other reviewers, instead of using specific reviewer pairs. Together with the limited number of reflective consensus meetings, this might have decreased inter-rater reliability. Also, the assessment by an expert team on transitional patient safety as the reference standard, as is often done in diagnostic studies, may be first choice.³⁰ However, this was not feasible in practice and the total of missed TSIs would increase because this expert team would have assessed all TSIs instead of only three items. Finally, correspondence of discharge and visits to the outpatient clinic was regularly missing. It was often unclear whether this was caused by failure in the data extraction, or the result of omission or faulty registration of correspondence in the hospital. Because of potential data extraction failure, the number of TSIs identified by both the reviewers and our research assistants could be overestimated. However, we did not register a missing letter as a TSI when any sign of a letter being sent was found in either the GP's or hospital specialist's medical record (such as a copy of the conclusion or new medication in the primary care record, or a note in the hospital record about sending a letter). In many of the reviewer-identified TSIs, the GP actually registered that no letter was received.

Implications

In order to become a valid and reliable measurement tool for TSIs, the medical record review method needs to be improved. Possibilities for improvement of medical record reviews are suggested in literature^{12 23–25 31}: making the transitional medical record database more complete, developing a more robust and standardised identification process, improving training and manual, and selecting a more homogeneous group of reviewers. Possibly, we could use one type of medical specialist, which would probably improve inter-rater reliability, but then would lead to missing specific types of TSI. When using a more standardised process, less specialised healthcare professionals could be used, such as research nurses, junior doctors or students in the final year of their medical study.

Although the measurement properties of the research methods should and can be improved, the research method does provide valuable information on the nature of TSIs. Besides using the medical record review for measurement of intervention effect, the identified TSIs could also be used for learning purposes: to raise awareness and influence improvement strategies, so TSIs can be prevented in the future.

CONCLUSION

Designing a valid assessment of transitional incidents in merged primary and secondary care medical records is challenging. In such assessment, clinicians indeed can identify TSIs that are clinically relevant and indicate room for improvement. Yet, the current identification method needs considerable adjustments to improve reliability and validity. This study serves as a starting point to optimise the identification process of TSIs.

Contributors All authors participated in the design of the study. MAVM, JMP and OJV acquired the data. MAVM, JMP, HFvS and OJV analysed and interpreted the data. MAVM, DLMZ, JMP, OJV and NjdW drafted the manuscript. HFvS and ML critically revised the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Dutch Ministry of Health, Welfare and Sports (VWS; grant number 320698), and Achmea Healthcare (grant number Z415).

Competing interests None declared.

Patient consent Not required.

Ethics approval The TIPP study was exempt from formal medical ethical approval according to Dutch law (METC numbers 13/142) and the TIPP medical record review method was approved as an addendum (METC addendum 13-142/C) by the Medical Ethical Committee, UMC Utrecht, the Netherlands.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Tam VC, Knowles SR, Cornish PL, *et al*. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. *CMAJ* 2005;173:510–5.
2. Forster AJ, Murff HJ, Peterson JF, *et al*. The incidence and severity of adverse events affecting patients after discharge from the hospital. *Ann Intern Med* 2003;138:161–7.
3. Moore C, Wisnivesky J, Williams S, *et al*. Medical errors related to discontinuity of care from an inpatient to an outpatient setting. *J Gen Intern Med* 2003;18:646–51.
4. Auerbach AD, Kripalani S, Vasilevskis EE, *et al*. Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA Intern Med* 2016;176:484–93.
5. Fertig A, Roland M, King H, *et al*. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *BMJ* 1993;307:1467–70.
6. Martinussen PE. Referral quality and the cooperation between hospital physicians and general practice: the role of physician and primary care factors. *Scand J Public Health* 2013;41:874–82.
7. Nash DB. "Lost in transition". *Prescriptions for Excellence in Health Care* 2010;10:1–4.
8. Coleman EA, Berenson RA. Lost in transition: challenges and opportunities for improving the quality of transitional care. *Ann Intern Med* 2004;141:533–6.
9. van Melle MA, Zwart DLM, de Bont AA, *et al*. Improving transitional patient safety: research protocol of the Transitional Incident Prevention Programme. *Safety in Health* 2015;1:13.
10. Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* 2013;22:273–7.
11. Weingart SN, Davis RB, Palmer RH, *et al*. Discrepancies between explicit and implicit review: physician and nurse assessments of complications and quality. *Health Serv Res* 2002;37:483–98.
12. Hanskamp-Sebregts M, Zegers M, Vincent C, *et al*. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open* 2016;6:e011078.
13. Vincent C, Burnett S, Carthey J. The measurement and monitoring of safety. The Health Foundation. 2013 http://www.health.org.uk/sites/health/files/TheMeasurementAndMonitoringOfSafety_fullversion.pdf (accessed 12 Jun 2017).
14. van Melle MA, Erkelens DC, van Stel HF, *et al*. Pilot study on identification of incidents in healthcare transitions and concordance between medical records and patient interview data. *BMJ Open* 2016;6:e011368.
15. COSMIN. COSMIN definitions of domains, measurement properties, and aspects of measurement properties. <http://www.cosmin.nl/images/upload/files/Tabel%20met%20definities-new.pdf> (accessed 12 Jun 2017).
16. Custodix. Secure and trustworthy re-use of EHR data for research. <https://www.custodix.com/>.
17. Baines RJ, Langelaan M, de Bruijne MC, *et al*. Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. *BMJ Qual Saf* 2013;22:290–8.
18. Zegers M, de Bruijne MC, Wagner C, *et al*. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;18:297–302.
19. van der Schaaf TW, Habraken MMP. PRISMA-Medical. A brief description. University of Technology, Eindhoven, The Netherlands. 2005 http://www.who.int/patientsafety/taxonomy/PRISMA_Medical.pdf (accessed 27 Jan 2017).
20. National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). Types of medication errors: NCC MERP index for categorizing medication errors. <http://nccmerp.org/sites/default/files/indexBW2001-06-12.pdf> (accessed 12 Jun 2017).
21. de Wet C, Bowie P. The preliminary development and testing of a global trigger tool to detect error and patient harm in primary-care records. *Postgrad Med J* 2009;85:176–80.
22. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 5th edn. Oxford UK: Oxford University Press, 2014.
23. Klopotoska JE, Wierenga PC, Stuijt CC, *et al*. Adverse drug events in older hospitalized patients: results and reliability of a comprehensive and structured identification strategy. *PLoS One* 2013;8:e71045.
24. Forster AJ, Taljaard M, Bennett C, *et al*. Reliability of the peer-review process for adverse event rating. *PLoS One* 2012;7:e41239.
25. Lilford R, Edwards A, Girling A, *et al*. Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy* 2007;12:173–80.
26. Göbel B, Zwart D, Hesselink G, *et al*. Stakeholder perspectives on handovers between hospital staff and general practitioners: an evaluation through the microsystems lens. *BMJ Qual Saf* 2012;21(Suppl 1):i106–13.
27. Hoffman C, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests. A systematic review. *JAMA Intern Med* 2017.
28. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;102:e93–e101.
29. van Beuzekom M, Boer F, Akerboom S, *et al*. Patient safety: latent risk factors. *Br J Anaesth* 2010;105:52–9.
30. Bertens LC, Broekhuizen BD, Naaktgeboren CA, *et al*. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med* 2013;10:e1001531.
31. Liddy C, Wiens M, Hogg W. Methods to achieve high interrater reliability in data collection from primary care medical records. *Ann Fam Med* 2011;9:57–62.