Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Diagnosis (Berl). Author manuscript; available in PMC 2015 June 19.

Published in final edited form as:

Diagnosis (Berl). 2015 February ; 2(1): 3–19. doi:10.1515/dx-2014-0047.

Clinical criteria to screen for inpatient diagnostic errors: a scoping review

Edna C. Shenvi, MD, MAS[*] and

Division of Biomedical Informatics, University of California, San Diego, 9500 Gilman Dr. MC 0728, La Jolla, CA 92093-0728, USA

Robert El-Kareh, MD, MS, MPH

Divisions of Biomedical Informatics and Hospital Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA, USA

Abstract

Diagnostic errors are common and costly, but difficult to detect. “Trigger” tools have promise to facilitate detection, but have not been applied specifically for inpatient diagnostic error. We performed a scoping review to collate all individual “trigger” criteria that have been developed or validated that may indicate that an inpatient diagnostic error has occurred. We searched three databases and screened 8568 titles and abstracts to ultimately include 33 articles. We also developed a conceptual framework of diagnostic error outcomes using real clinical scenarios, and used it to categorize the extracted criteria. Of the multiple criteria we found related to inpatient diagnostic error and amenable to automated detection, the most common were death, transfer to a higher level of care, arrest or “code”, and prolonged length of hospital stay. Several others, such as abrupt stoppage of multiple medications or change in procedure, may also be useful. Validation for general adverse event detection was done in 15 studies, but only one performed validation for diagnostic error specifically. Automated detection was used in only two studies. These criteria may be useful for developing diagnostic error detection tools.

Keywords

adverse event detection; diagnostic error detection; trigger tools

©2014, Edna C. Shenvi and Robert El-Kareh, published by De Gruyter.

[*]**Corresponding author: Edna C. Shenvi,** MD, MAS, Phone: 858-822-4931, eshenvi@ucsd.edu.

Author contributions: Conception and design of study: REK. Analysis and interpretation of data: ECS, REK. Drafting of the paper: ECS. Critical revision of paper for important intellectual content and final approval of the paper: REK, ECS. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Supplemental Material: The online version of this article (DOI: 10.1515/dx-2014-0047) offers supplementary material, available to authorized users.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

Diagnostic error can be defined as a wrong, missed, or delayed diagnosis [1] and is a cause of significant health-care harm that is largely preventable [2]. One estimate attributed diagnostic error for causing 40,000–80,000 deaths in the US annually in the inpatient setting alone, [3] and errors of diagnosis are the most common [4] and the most lethal [5] kind of professional liability claim. One in 20 US adults in the outpatient setting is estimated to be affected by a diagnostic error, [6] and about half of these errors are considered to be potentially harmful. While patient safety has become an increasingly high priority nationwide, diagnostic error has largely been overshadowed by efforts to reduce other kinds of harm, such as medication errors and nosocomial infections, and this may be due in part to the difficulty in measuring and analyzing diagnostic errors accurately.

Voluntary reporting and autopsies are some of the multiple possible approaches used to research diagnostic error, [7] but all have significant limitations. Retrospective chart review is often the best option, but this method is time-consuming and costly. Such review efforts have been facilitated by two-stage review processes, in which a nurse or other non-physician first reviews a chart for any among a list of screening criteria or “triggers”, such as an inpatient death or transfer to an intensive care unit (ICU), and those records that screen positive for a criterion are then reviewed by a physician to evaluate for the presence of an adverse event (AE). This method was first reported in the California Medical Insurance Feasibility Study (MIFS), [8, 9] adapted for the landmark Harvard Medical Practice Study (HMPS), [10–12] and similar studies in other countries [13–17], and influenced development of the “Global Trigger Tool” (GTT), [18] the most commonly employed such tool today.

None of these studies focused specifically on diagnostic errors, but usage of trigger tools has significant potential to improve the study of diagnostic error, [7] as this method can enrich the yield of charts reviewed and some can be applied with automated screening. Trigger tools for diagnostic error have been employed successfully in the outpatient setting using criteria such as an unscheduled hospitalization within 2 weeks of a primary care visit [19]. No such studies have been done specifically evaluating general triggers of diagnostic error in the inpatient setting. Study of diagnostic error has previously been focused in areas of high risk such as missed cancer in outpatient settings, and in the emergency department with high degrees of uncertainty and time pressure with undifferentiated patients. However, as these numerous studies on adverse events have demonstrated the harm and preventability of diagnostic error in hospitals, we sought to identify potential triggers that have been reported in research literature that could be used to screen for diagnosis-related errors in hospitalized adult patients, and specifically those that would be amenable to automated detection using data available in an electronic health record (EHR).

Methods

Developing a search strategy

We first compiled a test set of 12 articles by searching references in review articles and looking for related references through PubMed. From these articles we extracted keywords

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and medical subject headings (MeSH) terms, and evaluated iterations on our searches by success in retrieving the test set citations. Searches were then adapted for the syntax of Web of Science and CINAHL, with additional queries for those articles in PubMed not yet indexed with MeSH terms, and for those in CINAHL not yet similarly indexed. Keywords included “adverse events”, “diagnostic errors”, “detect”, and “identify”, and MeSH headings included “Medical Errors”, “Medical Audit”, and “Risk management/methods”. We limited our search to those published in the English language. A full description of the search strategy is available in the Supplementary Material.

Article selection

The final database searching in July 2013 of PubMed, Web of Science, and CINAHL retrieved 8861 references. Removing duplicates yielded 8558 unique citations. Both authors screened titles and abstracts, and identified 146 for full-text review. Disagreements were discussed between the two reviewers in order to come to a consensus. We also searched references cited in both research and review publications; this citation tracking yielded an additional five articles for inclusion, for a total of 33 included studies. Article selection is summarized in Figure 1.

We included only those articles that developed or validated criteria indicating that an error may have occurred. Some studies validated an entire tool, such as the GTT, but not individual criteria, so these were not included. Error-detection studies that referred to other articles for their methods or criteria used for their study were also excluded, although we searched these citations as well. Because our study focused on criteria applied to adult medical inpatients, we also excluded studies that were only applied to outpatients, emergency departments, pediatrics, or other specialties.

Framework of outcomes

We had a goal of categorizing potentially measurable trigger criteria that we would find in the literature search. As we did not find an appropriate system to use, we underwent an iterative process of creating a framework for this purpose. After preliminary discussion with clinical experts about the manifestations of diagnostic errors, we identified categories of “signals” related to patient status, clinical assessments (e.g., a diagnosis itself), and clinical actions (e.g., starting a treatment or making other management decisions). We used four appropriate clinical cases to help inform further development, and validated it with an additional four clinical case reports. As we reviewed the potential criteria published in the literature, we refined categorization in an iterative process. A depiction of this categorization is in Table 1.

Data extraction and analysis

One author [ES] primarily did the full-text review and data extraction, with guidance and revision by the other author [RE]. From each study, we extracted the research objectives and setting, presence of validation and automation, and also whether the authors reported that their methods detected “diagnostic error”, its incidence, and how such error was defined. A summary of included studies is in Table 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We extracted all record screening criteria from the included articles. Determination of whether or not such criteria were likely to be useful triggers for inpatient diagnostic error was done by comparison with our framework of potential outcomes. A summary of extracted criteria is in Table 3, grouped by category and ordered as presented in Table 1. An expanded version of all criteria reported is available in the Supplemental Material, Table 2.

Results

Criteria associated with diagnostic error

Significant clinical deterioration: death, arrest or code, and resultant clinical management—Error that results in harm would be those of highest priority to be prevented, which may manifest as some kind of clinical deterioration due to lack of management appropriate to the patient's true condition. The worst of such deterioration would result in a patient death, or near death, such as a cardiorespiratory arrest and the necessity for a “code team” or “rapid response team” to resuscitate the patient. Any inpatient “death” therefore would be a trigger criterion to prompt review for a diagnostic error, or for AEs in general. One or more of these three concepts (death, cardiorespiratory arrest, or medical emergency team response) was present in all studies. Some authors modified them to increase their specificity for error, such as by limiting deaths only to “unexpected death”, or “death unrelated to natural course of illness and differing from immediate expected outcome of patient management” [31]. One study [21] that was not attempting to measure incidence but to perform qualitative analysis of care prior to AEs only examined deaths subsequent to a code team call or ICU transfer. These criteria would all function differently compared with a categorical “death” criterion, although the concept is essentially the same.

Cardiorespiratory arrest was a separate criterion from death in most studies. To increase specificity of this criterion in the MIFS, reviewers were instructed to not count cardiorespiratory arrest as a positive screen in patients who were admitted for planned terminal care, [8] and Pavão [24] distinguishes this from “Death” by specifying only “reversed” cardiorespiratory arrest. It was linked with other “serious intervening event(s)” including deep vein thrombosis, pressure sore, and neurological events in two studies [32, 40]. Activation of a code team in response to such an arrest is a clinical action that would function very similarly as trigger, even though cardiorespiratory arrest itself is referring to the status of the patient. The three studies [22, 31, 37] that only listed an emergency team activation as a criterion (code, medical emergency, or rapid response team) did not have a separate cardiac arrest trigger. The GTT tool as developed in early studies [18, 27] has as one trigger “any code or arrest”, adding “rapid response team activation” in a later study, [22] but unlike others, does not have “Death” as a unique trigger separate from these clinical events except for intra- or post-operative death. Variations across hospitals for what clinical scenarios prompt such activations are likely to affect broad applicability of these criteria.

Transfer to an intensive care unit or other increased level of care was present in 29 of the studies. Several specified that such a transfer had to be unexpected for the record to screen positive. Names for units varied, with “intensive”, “semi-intensive”, “acute”, and “special” care units all being terms that were used, although only one [36] distinguished intensive and intermediate care transfer as two separate criteria. This criterion was broadened to any

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

"transfer to a higher level of care" in four studies [18, 22, 28, 35]. Three of these also included readmission to the ICU, [18, 22, 35] which could be a manifestation of deterioration in one missed disease process while another may have improved, but could be regarded as an overlap with any increased care acuity.

Other deterioration (as depicted in Table 3) may result in intubation after admission, new dialysis, or medical management changing to emergent surgery, resulting in an unexpected procedure or visit to the OR. We also grouped change of code status in this category, such as a patient being full code but then deteriorating to the point of receiving a "do not resuscitate" order. However, these four may not necessarily be the outcome of deterioration but may merely reflect a change in management plan, which is a category discussed below.

Unexpected time course of illness—Each condition has an expected range of potential courses, and some criteria aim to detect those courses that deviate from what was expected. Diagnostic error may manifest after discharge, such as lack of improvement necessitating return to healthcare or readmission within a certain time period. Many studies used readmission within a threshold number of days, or specifically a readmission because of the care provided in the previous admission [33].

Diagnostic error may also prolong the hospital course, with or without deterioration or changes in clinical management. This concept was present in 10 of the studies, although quantified in different ways: using threshold values (e.g., length of stay [LOS] >35, 21, or 10 days), comparison with an "expected" duration at admission, or comparison with average duration for diagnosis-related group (DRG). The HMPS [10] used different percentiles for DRG based on patient age. Craddick's Medical Management Analysis (MMA) program [48] was a system intended to be tailored to each hospital, and this system left it to each institution's discretion to set its own threshold of either LOS or a percentile. Specific thresholds are less broadly applicable across geographic and care settings, as Kobayashi et al report [30] that the average length of hospital stay in Japan in 2004 was 22.2 days, much longer than in other countries. This makes threshold values for length of stay unable to be applied in Japan as they are elsewhere. The MIFS group [8] attempted to improve the specificity of this criterion by instructing reviewers to exclude those prolongations of hospitalization that were only for administrative or social reasons.

Change of management team—Recognition of a diagnostic error could result in a patient transfer to another hospital, which was a criterion included in 23 studies. Two [8, 24] include exceptions for those transfers that are for exams or procedures unavailable at the first hospital and those that are mandatory for administrative reasons. This may be of questionable utility in urban tertiary care academic medical centers, but useful for quality review purposes in smaller hospitals, in which it may occur that a patient's failure to improve results in a decision to procure assistance from a larger or more specialized institution.

Correction of error could also result in a patient transfer to another service, such as from general medicine to cardiology, without a concomitant increase in acuity. Resar and colleagues [35] in the ICU trigger tool included change of physician in charge as a criterion

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for detection of potential error. While physician changes happen for multiple reasons besides diagnostic error, construction of a tool to detect change of management team may be a useful screen.

Change of specific management plan—Responding to or correcting a diagnostic error could result in a change in the specific treatment plan for a patient. “Abrupt medication stop” was a criterion in five studies; one [23] specified this for an enterprise data warehouse query as “discontinuation of 4 of more medications in a 6 h period >48 h after admission and at least 24 h prior to discharge”.

Changes in plan regarding a surgery or other procedure could manifest in three different ways, and all were present in at least one study. A patient with a wrong diagnosis could be booked for surgery and then changed to medical management (i.e., cancellation of a planned procedure), or have a pre-operative diagnosis and planned procedure that differs significantly from a post-operative diagnosis and actual procedure performed (i.e., change in procedure). Alternatively, an unplanned visit to the operating room or other procedural facility, as discussed under deterioration, could also occur after a patient receives medical therapy only to eventually undergo operative management.

Change in diagnosis itself—Recognition of an incorrect diagnosis could reasonably result in a modification of the patient's primary or secondary problems in the medical record. The only criterion we encountered that dealt with a diagnosis was a pathology result either normal or unrelated to the previous diagnosis, [18, 27] which we grouped with criteria related to procedures since this would be the mode of acquiring a specimen, and the comparison would be made to a pre-procedure diagnosis. Other criteria related to diagnoses, as listed in our “Indicators of Clinical assessment” column in the framework, were not present in these studies. Substantial change in the problem list might be a useful screen for diagnostic error, although the performance and potential for automation would vary by EHR design and on how well the problem list is maintained.

Diagnostic uncertainty—Identification of cases in which the correct diagnosis was unclear to the medical team would be useful for review for patient safety and educational purposes. Direct measurement of instances of diagnostic uncertainty would be difficult; however, there are possibly ways to detect indirect manifestations. We found only one published criterion in this category. Resar's ICU trigger tool [35] used the criterion of multiple consultations, using the threshold of three or more. Requesting input from multiple specialties could be a manifestation of either diagnostic uncertainty or delay.

Other criteria reported likely to be associated with diagnostic error

Other criteria we encountered may be associated with diagnostic error, but are unlikely to be amenable to automated detection or available in an EHR. Two studies [31, 32] used delay or error in diagnosis as a screening criterion itself, which would require significant clinician interpretation in order to use. Others, such as patient and provider dissatisfaction, litigation, or ethics board referrals, are not likely to be in an EHR.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Criteria for non-diagnostic error

The criteria we regarded as not associated with diagnostic error were those developed to detect other types of adverse events, such as those specific for nosocomial infections (positive blood or urine cultures), adverse drug events (administration of "rescue" medications, like naloxone or vitamin K), and procedural complications (return to the operating room or injury of an organ during a procedure). Additional triggers were for in-hospital falls, venous thromboembolic events, strokes, pressure ulcers, and bleeding. Since the majority of these studies employed manual screening, many also had a vague "catch-all" category that is not translatable to an electronic trigger; two such criteria are "any other undesirable outcome not covered above" [10] and "other finding on chart review suggestive of an adverse event" [44]. A full list of these criteria we considered not primarily associated with electronic detection of diagnostic error, as well as example studies in which each was used, is available in the Supplementary Material, Table 2.

Diagnostic error, validation and automation

Thirteen of the studies reported the proportion of adverse events involving diagnostic error, while the other studies had varied research objectives, mostly determining severity and preventability of adverse events rather than other categorizations. In the studies that reported results for diagnostic errors, the proportion of AEs involving diagnostic errors was as low as 5.1% in one study [33] to 67.5% in another [37] (discussed below). These proportions are difficult to compare given variation in the definitions of their study populations (denominators).

Fifteen studies reported validation metrics for their criteria, as positive predictive values or odds ratios for individual criteria, or kappa coefficients for inter-rater agreement during manual review. However, these are validation metrics for AEs in general and not specific to diagnosis. Only one study [37] provides statistics for which validation in detecting diagnostic error can be inferred. This study used only the single criterion of medical emergency team referral, and found that 31.3% were associated with medical errors, of which 67.5% were determined to be diagnostic, although it used a broader definition of diagnostic error that includes failure to perform a test or act on known results (as in Table 2). Given the variability of definitions of error and care practices, validation metrics cannot be aggregated across the studies at this time.

Only two studies used automated methods: one compared automated with traditional manual triggers [23] and another electronically screened discharge summary text [41]. These did not report proportion of errors that were diagnostic in nature. Bates et al. [46] described that many AEs could be detected by computer systems even with low levels of sophistication, but even in the two decades since publication, this has rarely been done or reported in research literature for AE detection that includes diagnostic error.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Discussion

Criteria in outcomes framework and potential for automated detection

Five of the criteria used in much of the original work in AE detection can be associated with inpatient diagnostic error: death, cardiorespiratory arrest (or code), transfer to a higher level of care, long length of stay, and transfer to another hospital. Subsequent trigger tools added other indicators: calling multiple consults, change of physician in charge, changes in procedure, abrupt medication stop, and discordance between diagnosis and pathology result. We also included change of code status, intubation, and new dialysis as those with reasonable association with diagnostic error.

In current EHR systems, it may be possible to use multiple changes in active diagnoses as triggers for detecting diagnostic error, as we depict in our model. A change of organ system for a primary diagnosis, such as pancreatitis to myocardial infarction, for example, could be a reasonable trigger, as well as addition of or significant changes to secondary diagnoses. We also considered some indicators of diagnostic uncertainty may be useful, such as a symptom or finding-based primary diagnosis rather than a true diagnosis, and multiple diagnostic procedures at a certain point in hospitalization, similar to multiple consults. However, we did not find their use reported in either manual or automated methods.

None of these criteria is specific to diagnostic error, and many may be difficult to translate into an automated query. A delay in diagnosis or initiation of effective treatment, which are triggers in three studies [31, 32, 40], would need to be more explicitly defined to use in automated detection. One possible way to concretize the concept of treatment delay could be usage of a threshold number of hospital day for first surgery, similar to using thresholds for long length of stay, such as first major procedure on or after hospital day 3. We encountered this concept in the Complications Screening Program, [49] although it was only used for risk stratification and was not a trigger in itself, therefore not included in our table. A 2-day delay from admission to surgery, however, might also be a manifestation of delay in diagnosis, and may warrant further evaluation as a potential screen for diagnostic error.

Trigger tools can be used for both retrospective detection of errors, or real-time activation of some kind of intervention to prevent imminent or potential harm. Some criteria are by nature exclusively retrospective, such as a patient death, but others may indicate uncertainty or error, which in real time could possibly be used to trigger special attention to a particular clinical case. The studies we included focused on retrospective error detection, but some may have potential for real-time prevention of harm as well, such as multiple consultations or change in multiple medications, that could prompt additional attention to a particular case. We anticipate that new trigger criteria for both purposes could be developed in the future, and we would hope that the conceptual framework we developed on the manifestations of diagnostic error would provide a structure around which new criteria and formalized understanding of this area can be organized.

Identifying those triggers that will be the most useful for detecting diagnostic error will require much further study. Positive predictive values, if reported in studies as validation metrics, were for general error detection and not specific for diagnostic error. Eleven of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

these studies reported proportions of errors that were diagnostic, but definitions of what constituted this type of error varied. Both Mills [8] and Soop [29] further sub-categorized wrong and delayed diagnoses while some studies grouped these two together; however, a few of the definitions may be more broad than generally used. The inclusion of failure to order tests or misinterpretation of results used in two studies [37, 38], and the grouping of diagnostic with treatment delay in another [36] demonstrates the variability in defining diagnostic error, but also illustrates some of the multiple ways it can manifest. For trigger criteria to be more useful, it would be necessary to establish a consensus on the definition of error. Such a consensus would allow comparison of the performance of different criteria across multiple settings.

Limitations

Our objective was to compile available trigger criteria reported in the literature that may be useful in development of screening tools to detect inpatient diagnostic error. As this was a difficult concept to clearly and exhaustively query in research databases, even our iterative approach in finalizing a search strategy gave us a rather low yield of articles for the number of original citations retrieved. While we had two reviewers screening citations, only one primarily did the extraction. We also only reviewed English language publications, and only research on adult medical patients. It is possible that these limitations may have resulted in missing a few more such screening criteria, although many of the studies we found repeated similar criteria and cited similar sources, causing us to conclude that our citation tracking was sufficient and a more extensive literature search was of low likelihood to yield many more new criteria. Additionally, our framework of outcomes was based on a small corpus of case reports; further validation with more cases is warranted.

Determining association of specific criteria with diagnostic error was also not straightforward. While it could be argued that any AE may have some component of diagnosis or assessment involved, we focused on criteria for which the primary measured item was potentially directly related to a diagnostic assessment. Both our conceptual framework and our compiled criteria are of a global or convergent picture of diagnostic error, rather than a comprehensive tool for detecting all possible diagnoses that could be missed or wrong.

Conclusions

Inpatient diagnostic errors may be more easily detected and studied using available “triggers” to facilitate chart review. We have identified several such criteria that could be used to develop automated screening tools for detection of diagnostic errors using available data within electronic health records. In addition, we developed a preliminary conceptual framework of outcomes to formalize the manifestations of inpatient diagnostic error that we hope will be helpful and expanded in the future with further study in this area. We identified some additional criteria that may also be useful, but have not been used in either manual or automated methods to date. Validation of these criteria is needed to identify those that will provide the most effective screen for inpatient diagnostic error.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the guidance and assistance in the development of our search strategy by Mary Wickline, Nancy Stimson, and Penny Coppernoll-Blach, who are all librarians at the University of California, San Diego Library.

Financial support: Dr. Shenvi is supported by the National Library of Medicine training grant T15LM011271, San Diego Biomedical Informatics Education and Research. Dr. El-Kareh is supported by K22LM011435-02, a career development award from the National Library of Medicine.

References

1. Graber M. Diagnostic errors in medicine: a case of neglect. Jt Comm J Qual Patient Saf. 2005; 31:106–13. [PubMed: 15791770]
2. Zwaan L, de Bruijne M, Wagner C, Thijs A, Smits M, van der Wal G, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. Arch Intern Med. 2010; 170:1015–21. [PubMed: 20585065]
3. Leape L, Berwick D, Bates D. Counting deaths due to medical errors-Reply. J Am Med Asssoc. 2002; 288:2404–5.
4. Mangalmurti SS, Harold JG, Parikh PD, Flannery FT, Oetgen WJ. Characteristics of medical professional liability claims against internists. J Am Med Assoc Intern Med. 2014; 174:993–5.
5. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. Br Med J Qual Saf. 2013; 22:672–80.
6. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. Br Med J Qual Saf. 2014; 23:727–31.
7. Graber ML. The incidence of diagnostic error in medicine. Br Med J Qual Saf. 2013; 22(Suppl 2):ii21–ii7.
8. Mills, DH., editor. Sponsored jointly by California Medical Association and California Hospital Association. Sutter Publications, Inc.; San Francisco, CA: 1977. Report on the Medical Insurance Feasibility Study..
9. Mills DH. Medical insurance feasibility study. A technical summary. West J Med. 1978; 128:360–5. [PubMed: 664637]
10. Hiatt HH, Barnes BA, Brennan TA, Laird NM, Lawthers AG, Leape LL, et al. A study of medical injury and medical malpractice. N Engl J Med. 1989; 321:480–4. [PubMed: 2761585]
11. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med. 1991; 324:370–6. [PubMed: 1987460]
12. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991; 324:377–84. [PubMed: 1824793]
13. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. Med J Aust. 1995; 163:458–71. [PubMed: 7476634]
14. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. Br Med J. 2001; 322:517–9. [PubMed: 11230064]
15. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals I: occurrence and impact. N Z Med J. 2002; 115:U271. [PubMed: 12552260]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

16. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. Can Med Assoc J. 2004; 170:1678–86. [PubMed: 15159366]
17. Zegers M, de Bruijne MC, Wagner C, Hoonhout LH, Waaijman R, Smits M, 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. [PubMed: 19651935]
18. Classen DC, Lloyd RC, Provost L, Griffin FA, Resar R. Development and evaluation of the institute for healthcare improvement global trigger tool. J Patient Safety. 2008; 4:169–77.
19. Singh H, Giardina TD, Forjuoh SN, Reis MD, Kosmach S, Khan MM, et al. Electronic health record-based surveillance of diagnostic errors in primary care. Br Med J Qual Saf. 2012; 21:93–100.
20. Cihangir S, Borghans I, Hekkert K, Muller H, Westert G, Kool RB. A pilot study on record reviewing with a priori patient selection. BMJ Open. 2013; 3:e003034. pii.
21. De Meester K, Van Bogaert P, Clarke SP, Bossaert L. In- hospital mortality after serious adverse events on medical and surgical nursing units: a mixed methods study. J Clin Nurs. 2013; 22:2308–17. [PubMed: 22827923]
22. Hwang JI, Chin HJ, Chang YS. Characteristics associated with the occurrence of adverse events: a retrospective medical record review using the Global Trigger Tool in a fully digitalized tertiary teaching hospital in Korea. J Eval Clin Pract. 2014; 20:27–35. [PubMed: 23890097]
23. O'Leary KJ, Devisetty VK, Patel AR, Malkenson D, Sama P, Thompson WK, et al. Comparison of traditional trigger tool to data warehouse based screening for identifying hospital adverse events. Br Med J Qual Saf. 2013; 22:130–8.
24. Pavão AL, Camacho LA, Martins M, Mendes W, Travassos C. Reliability and accuracy of the screening for adverse events in Brazilian hospitals. Int J Qual Health Care. 2012; 24:532–7. [PubMed: 22879373]
25. Wilson RM, Michel P, Olsen S, Gibberd RW, Vincent C, El-Assady R, et al. Patient safety in developing countries: retrospective estimation of scale and nature of harm to patients in hospital. Br Med J. 2012; 344:e832. [PubMed: 22416061]
26. Letaief M, El Mhamdi S, El-Asady R, Siddiqi S, Abdullatif A. Adverse events in a Tunisian hospital: results of a retrospective cohort study. Int J Qual Health Care. 2010; 22:380–5. [PubMed: 20685729]
27. Naessens JM, O'Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. Int J Qual Health Care. 2010; 22:266–74. [PubMed: 20534607]
28. Cappuccio FP, Bakewell A, Taggart FM, Ward G, Ji C, Sullivan JP, et al. Implementing a 48 h EWTD-compliant rota for junior doctors in the UK does not compromise patients' safety: assessor-blind pilot comparison. Q J Med. 2009; 102:271–82.
29. Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. Int J Qual Health Care. 2009; 21:285–91. [PubMed: 19556405]
30. Kobayashi M, Ikeda S, Kitazawa N, Sakai H. Validity of retrospective review of medical records as a means of identifying adverse events: comparison between medical records and accident reports. J Eval Clin Pract. 2008; 14:126–30. [PubMed: 18211655]
31. Mitchell IA, Antoniou B, Gosper JL, Mollett J, Hurwitz MD, Bessell TL. A robust clinical review process: the catalyst for clinical governance in an Australian tertiary hospital. Med J Aust. 2008; 189:451–5. [PubMed: 18928440]
32. Williams DJ, Olsen S, Crichton W, Witte K, Flin R, Ingram J, et al. Detection of adverse events in a Scottish hospital using a consensus-based methodology. Scott Med J. 2008; 53:26–30. [PubMed: 19051661]
33. Sari AB, Sheldon TA, Cracknell A, Turnbull A, Dobson Y, Grant C, et al. Extent, nature and consequences of adverse events: results of a retrospective casenote review in a large NHS hospital. Qual Saf Health Care. 2007; 16:434–9. [PubMed: 18055887]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

34. Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, Waaijman R, van der Wal G. Design of a retrospective patient record study on the occurrence of adverse events among patients in Dutch hospitals. BMC Health Serv Res. 2007; 7:27. [PubMed: 17319971]
35. Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf. 2006; 32:585–90. [PubMed: 17066996]
36. Herrera-Kiengelher L, Chi-Lem G, Báez-Saldaña R, Torre-Bouscoulet L, Regalado-Pineda J, López-Cervantes M, et al. Frequency and correlates of adverse events in a respiratory diseases hospital in Mexico city. Chest. 2005; 128:3900–5. [PubMed: 16354861]
37. Braithwaite R, DeVita M, Mahidhara R, Simmons R, Stuart S, Foraida M. Use of medical emergency team (MET) responses to detect medical errors. Qual Saf Health Care. 2004; 13:255–9. [PubMed: 15289627]
38. Forster AJ, Asmis TR, Clark HD, Al Saied G, Code CC, Caughey SC, et al. Ottawa Hospital Patient Safety Study: incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. CMAJ. 2004; 170:1235–40. [PubMed: 15078845]
39. Michel P, Quenon JL, de Sarasqueta AM, Scemama O. Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. Br Med J. 2004; 328:199. [PubMed: 14739187]
40. Chapman EJ, Hewish M, Logan S, Lee N, Mitchell P, Neale G. Detection of critical incidents in hospital practice: a preliminary feasibility study. Clinical Governance Bulletin. 2003; 4:8–9.
41. Murff H, Forster A, Peterson J, Fiskio J, Heiman H, Bates D. Electronically screening discharge summaries for adverse medical events. J Am Med Inform Assoc. 2003; 10:339–50. [PubMed: 12668691]
42. Wolff AM, Bourke J, Campbell IA, Leembruggen DW. Detecting and reducing hospital adverse events: outcomes of the Wimmera clinical risk management program. Med J Aust. 2001; 174:621–5. [PubMed: 11480681]
43. Thomas E, Studdert D, Burstin H, Orav E, Zeena T, Williams E, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000; 38:261–71. [PubMed: 10718351]
44. Bates DW, O'Neil AC, Petersen LA, Lee TH, Brennan TA. Evaluation of screening criteria for adverse events in medical patients. Med Care. 1995; 33:452–62. [PubMed: 7739271]
45. Wolff AM. Limited adverse occurrence screening: an effective and efficient method of medical quality control. J Qual Clin Pract. 1995; 15:221–33. [PubMed: 8719061]
46. Bates DW, O'Neil AC, Boyle D, Teich J, Chertow GM, Komaroff AL, et al. Potential identifiability and preventability of adverse events using information systems. J Am Med Inform Assoc. 1994; 1:404–11. [PubMed: 7850564]
47. O'Neil AC, Petersen LA, Cook EF, Bates DW, Lee TH, Brennan TA. Physician reporting compared with medical-record review to identify adverse medical events. Ann Intern Med. 1993; 119:370–6. [PubMed: 8338290]
48. Craddick, JW.; Bader, B. Medical management analysis: a systematic approach to quality assurance and risk management. Joyce W. Craddick; Auburn, California: 1983.
49. Iezzoni LI, Daley J, Heeren T, Foley SM, Fisher ES, Duncan C, et al. Identifying complications of care using administrative data. Med Care. 1994; 32:700–15. [PubMed: 8028405]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

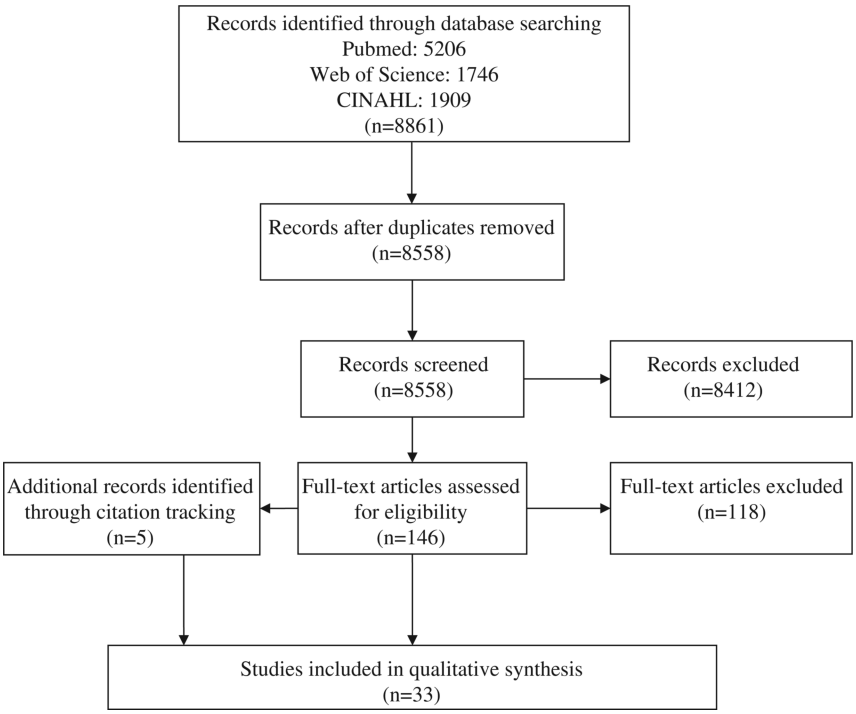

Figure 1.
Article selection.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Framework of outcomes of inpatient diagnostic error.

Category of manifestations of error	Indicators of patient status	Indicators of clinical assessment	Indicators of clinical management
Patient deterioration	–Death* – Cardiorespiratory arrest*		–Call code, rapid response or medical emergency team* –Transfer to higher level of care* –Unexpected emergent treatment (e.g. intubation, dialysis, procedure)*
Unexpected time course of illness	– Ongoing symptoms necessitating representation to healthcare system*		–Prolonged hospitalization* –Shorter than expected length of stay –Treatment extended beyond normal duration –Readmission for related symptoms or condition*
Change of management plan (recognition of error)		–Change in primary dx –Add, change, or remove secondary dx –Discrepancy between reason for admission and subsequent primary dx	–Transfer to another hospital* –Change in primary service (Change of physician in charge*) –Abrupt starting or stopping multiple medications* –Change in procedure dx or type*
Diagnostic uncertainty		– Symptom or findingbased primary dx	–Multiple consultations* –Multiple diagnostic procedures

Note that these categories are not meant to be mutually exclusive, in that any single case of diagnostic error can ultimately manifest in any number of these potential outcomes. Those marked with an asterisk (*) were those we found as screening criteria in our review.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Characteristics of included studies.

First author	Year	Number of criteria	Study intent	Record selection, timing, and geographic setting	Number of records applied	Detection of diagnostic error reported?	Validation metrics?	Automated?
Cihangir et al. [20]	2013	1+GTT	Validate unexpectedly long length of stay (UL-LOS) as a screening tool to use with the GTT to prompt record review for AE	Colon cancer admissions (excluding palliative patients and those who died in the hospital), in 2009: 1 general hospital (Tergooiziekenhuizen) in the Netherlands	129	No: only determined severity	Yes: proportions	No
De Meester et al. [21]	2013	2	Assess nursing care prior to serious adverse events resulting in patient death, for extent of preventability	All admissions, March 1 2007–November 15 2007: 1 teaching hospital in Belgium	14,106	No: only analyzed nursing care and preventability	No	No
Hwang et al. [22]	2014	53	Examine performance of GTT and investigate characteristics associated with AEs	Random 30 charts per week, January–June 2011: 1 tertiary teaching hospital in Korea	629	No: categorization of AEs by organ system, preventability, and harm	Yes: PPVs in Table 2	No
O'Leary et al. [23]	2013	51 based on 33	Compare 33 traditional trigger queries (based on HMPS) to an enterprise data warehouse screening method, made of 51 automated triggers based on manual triggers.	Random 250 general medicine admissions (excluding those admitted only for observation or patients of the study's authors), September 1 2009–August 31 2010: Northwestern Hospital, Chicago, Illinois	250	No: does report categories of ADEs, HAIs, VTE, fall, etc.	Yes: PPVs in Table 4	Yes
Pavão et al. [24]	2012	19	Compare screening by nurses to residents for AE detection; validation of metrics from previous study done to detect AEs	Random sample of 1103 adult non-psychiatric admissions, in 2003: 3 general teaching hospitals in Rio de Janeiro, Brazil	242	No: no categorization of AEs given	Yes: percentage by reviewer type, and kappa agreement, Sn, Sp, PPV, and NPV for top 6	No
Wilson et al. [25]	2012	18	Assess frequency and nature of AEs in developing or transitional economies	Random sample of at least 450 records per site, in 2005: convenience sample of 26 hospitals from Egypt, Jordan, Kenya, Morocco, Tunisia, Sudan, South Africa, and Yemen	15,548	Yes: 19.1% (12–41%) defined as failure to make a diagnosis or to do so in a timely manner or the failure to make a correct diagnosis from provided information	No	No
Letaief et al. [26]	2010	18	Estimate incidence and nature of AEs in hospital	Stratified random sampling of admissions, in 2005: one university hospital in Tunisia	620	Yes: 12.9% of AEs	Yes for top 5: PPV and RR	No

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First author	Year	Number of criteria	Study intent	Record selection, timing, and geographic setting	Number of records applied	Detection of diagnostic error reported?	Validation metrics?	Automated?
Naessens et al. [27]	2010	55	Assess performance of GTT and inter-rater reliability	10 random adult inpatients every 2 weeks, August 2005–September 2007 or August 2006–September 2007 (different sites): 3 Mayo clinic campuses in Minnesota, Arizona, and Florida	1138	No: categorization only by level of harm	Yes: kappa for inter-rater agreement, PPVs	No
Cappuccio et al. [28]	2009	28	Determine if AEs more or less common after decreasing work hours of junior doctors	All records, 12-week period in 2007: 1 hospital in Coventry, UK	1581	No: categorized by severity and preventability	No	No
Soop et al. [29]	2009	18	Determine incidence and nature of AEs in Sweden	National sample of 1.2 million admissions, October 2003–September 2004: 28 hospitals in Sweden	1967	Yes: 11.3% of AEs, also subcategorized as wrong (0.8%), failed or delayed (7.1%), and incomplete diagnosis (0.4%), and diagnostic procedure AEs (2.9%)	Yes: Sn and PPV	No
Classen et al. [18]	2008	55 in 6 modules	Develop GTT and do preliminary testing, calculate kappas between reviewers	Sample records from standardized training systems from Midwestern United States	15 train+50 test	No: only determination of severity	No	No
Kobayashi et al. [30]	2008	18	Determine if record review is a valid means of detecting AEs in accident reports	Random 100 cases+100 accident reports, in 2002: one hospital in Japan	200	No: no formal categorization of all AEs detected	No	No
Mitchell et al. [31]	2008	7 flags, 17 other triggers	Assessing implementation of a Clinical Review Committee for hospital governance	All records, September 1 2002–June 30 2006: Canberra Hospital, Australia	5925	No: categorized by severity and associated care systems	No	No
Williams et al. [32]	2008	15	Determine rate and nature of inpatient AEs (medical, surgical, obstetric units)	150 consecutive admissions >24 h per unit (medical and surgical units of 1 hospital and obstetric admissions at maternity hospital), 1 month starting July 1 2004: Aberdeen, Scotland	354	No: categorized by severity and associated care processes	No	No
Sari et al. [33]	2007	18	Determine incidence and nature of AEs and evaluate review methods	Random sample of admissions, January–May 2004: one large hospital in England	1006	Yes: 5.1% of AEs	Yes: ORs and kappa agreement	No

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First author	Year	Number of criteria	Study intent	Record selection, timing, and geographic setting	Number of records applied	Detection of diagnostic error reported?	Validation metrics?	Automated?
Zegers et al. [34]	2007	18	Design a method to study the occurrence of AEs in Dutch hospitals	Discharges (excluding obstetric and psychiatric admissions) or deaths, in 2004: 21 hospitals in the Netherlands	8400	No: classification only on preventability and degree of management causation	No	No
Resar et al. [35]	2006	23	Identify AEs in intensive care units	10 charts per month, 80–500 per organization, in 2001–2004: 62 ICUs in 54 hospitals in Institute for Healthcare Improvement collaborative	1294	No: classification only by severity	Yes: PPVs for 6 of the 23	No
Herrera-Kiengelher et al. [36]	2005	15	Estimate the frequency of severe AEs in hospital stays and correlates for respiratory disease	Admissions from year 2001: one hospital dedicated to respiratory diseases in Mexico City, Mexico	4555	Yes: “delayed diagnosis and/or treatment,” 41/415 AEs (9.8%)	Yes: PPV	No
Baker et al. [16]	2004	18	Determine the incidence of AEs in Canada	Random sample of charts, in 2000: 4 hospitals in each of 5 provinces in Canada	3745	Yes: 38/360 AEs (10.6%), discussed as “failure to carry out necessary diagnosis”	No	No
Braithwaite et al. [37]	2004	1	Determine if medical emergency team activation is good indicator of error	All emergency response team activations, May–December 2008: 3 connected hospitals at University of Pittsburgh, Pennsylvania	364	Yes: 67.5% of AEs (errors that resulted from improper or delayed diagnosis, failure to employ indicated tests, or failure to act on the results of monitoring or testing)	Yes	No
Forster et al. [38]	2004	16	Assess timing within hospitalization of AEs	Random 502 admissions over 1 year (exact year N/A): Ottawa Hospital, Ottawa, Canada	502	Yes: 9% of AEs (an indicated test was not ordered or a significant test result was misinterpreted)	No	No
Michel et al. [39]	2004	17	Compare prospective, cross sectional, and retrospective methods of detecting AEs	(Timing N/A) 37 wards in 3 public and 4 private hospitals in Aquitaine, France	778	No: categorized only by preventability	No	No
Chapman et al. [40]	2003	14+optional 1	Determine feasibility of reviewing case records to identify critical incidents in ‘real time’	Consecutive patients over 2weeks (exact timing N/A): 3wards in one hospital in London, United Kingdom	76	No: no categorization of AEs given	No	No
Murff et al. [41]	2003	11 concepts, 95 trigger words	Determine if electronic screening of discharge summaries can detect AEs	Random admissions, January 1–June 30 2000: Brigham and Women's Hospital, Boston, Massachusetts	837	Yes: 7.4% of AEs	Yes: PPVs of trigger words	yes

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First author	Year	Number of criteria	Study intent	Record selection, timing, and geographic setting	Number of records applied	Detection of diagnostic error reported?	Validation metrics?	Automated?
Wolff et al.[a] [42]	2001	8	Determine if a risk management program of detecting and acting on AEs can reduce the rate of harm	All admissions, July 1991–September 1999: one rural hospital in Wimmera region of Victoria, Australia	49,834	No: categorized by likelihood and consequence (severity)	No	No
Thomas et al. [43]	2000	18	Determine incidence and types of AEs in Utah and Colorado, to assess generalizability of findings of other studies	Random non-psychiatric discharges, in 1992: sample of hospitals in Utah and Colorado	15,000	Yes: incorrect or delayed diagnosis, 6.9% of AEs	No	No
Wilson et al. [13]	1995	18	Estimate patient injury in Australian hospitals	Random admissions, in 1992: 28 hospitals in New South Wales and South Australia	14,179	Yes: 13.6% of AEs. Defined as AE arising from a delayed or wrong diagnosis	Yes: OR	No
Bates et al. [44]	1995	15	Evaluate screening criteria for AEs	Consecutive admissions to 1 medical service, November 1990–March 1991: Brigham and Women's Hospital, Boston, Massachusetts	3137	No: categorization only by severity and preventability	Yes: Sn, Sp, PPV, OR	No
Wolff [45]	1995	8	Determine rate of AEs, and assess effectiveness of limiting review to those positive for 8 screening criteria	All discharges, July 1 1991–June 30 1992: Wimmera Base Hospital in Horsham, Victoria, Australia	5115	No: only assessed causation and severity	Yes: PPV	No
Bates et al. [46]	1994	13	Evaluate ability of computerized systems to identify AEs detected by manual review	All medical admissions, November 1990–March 1991: Brigham and Women's Hospital, Boston, Massachusetts	3146	No: AEs discussed in a previous paper[47] only as related to procedure, medication, or other therapy	No	No
Hiatt et al. [10]	1989	18	Measure incidence of AEs and determination of negligence	Random records, in 1984: 51 hospitals in state of New York	30,121	Yes: in Leape [12] 8.1% of AEs, defined as improper or delayed diagnosis	No	No
Craddick and Bader [48]	1983	23	Description of Medical Management Analysis system for individual hospitals to use for their own quality assurance	Tailored to each hospital	N/A	N/A	No	No
California Medical Association and California Hospital Association, Mills [8]	1977	20	Determine the cost feasibility of almost any non-fault compensation system; report the type, frequency and severity of patient disabilities caused by health care	Year 1974: representative sample of 23 nonfederal, short-term general hospitals in California	20,864	Yes: misdiagnosis 10% and nondiagnosis (failure to diagnose) 13% of PCEs	No	No

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First author	Year	Number of criteria	Study intent	Record selection, timing, and geographic setting	Number of records applied	Detection of diagnostic error reported?	Validation metrics?	Automated?
			management (“potentially compensable event” or PCE)					
			management (“potentially compensable event” or PCE)					
			management (“potentially compensable event” or PCE)					
			management (“potentially compensable event” or PCE)					

ADE, adverse drug event; AE, adverse event; GTT, global trigger tool; HAI, healthcare associated infection; HMPS, Harvard medical practice study; N/A, not applicable; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value; RR, relative risk; Sn sensitivity; Sp, specificity; VTE, venous thromboembolism.

[a]Wolff [42] includes two studies, one performed on inpatients and another in the emergency department. Only the numbers for the inpatient study are reported here. The emergency department study used a separate set of criteria, such as length of stay >6 h.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Signals of potential inpatient diagnostic error.

Criteria amenable to automated detection			
Category	**Unique concepts**	**Specific wording**	**Study(ies) used**
Patient deterioration: Indicators of patient status	Death	Death	Mills [8], Hiatt et al. [10], Brennan et al. [11], De Meester et al. [21], O'Leary et al. [23], Pavão et al. [24], Cappuccio et al. [28], Mitchell et al. [31], Williams et al. [32], Resar et al. [35], Herrera-Kiengelher et al. [36], Chapman et al. [40], Murff et al. [41], Wolff et al. [42], Thomas et al. [43], Bates et al. [44], Wolff [45], Bates et al. [46], Craddick and Bader [48]
		Unexpected death	Wilson et al. [13], Baker et al. [16], Wilson et al. [25], Letaief et al. [26], Soop et al. [29], Kobayashi et al. [30], Mitchell et al. [31], Sari et al. [33], Zegers et al. [34], Forster et al. [38], Michel et al. [39]
	Cardiac/respiratory arrest	Cardiorespiratory arrest	Mills [8], Hiatt et al. [10], Brennan et al. [11], Wilson et al. [13], Baker et al. [16], Classen et al. [18], O'Leary et al. [23], Wilson et al. [25], Letaief et al. [26], Naessens et al. [27], Cappuccio et al. [28], Soop et al. [29], Kobayashi et al. [30], Williams et al. [32], Sari et al. [33], Zegers et al. [34], Forster et al. [38], Michel et al. [39], Chapman et al. [40], Murff et al. [41], Wolff et al. [42], Thomas et al. [43], Bates et al. [44], Wolff [45], Bates et al. [46], Craddick and Bader [48]
Patient deterioration: Indicators of clinical management	Activation of teams responding to acute patient decompensation	Code, rapid response, or medical emergency team activation	Classen et al. [18], De Meester et al. [21], Hwang et al. [22], Naessens et al. [27], Mitchell et al. [31], Resar et al. [35], Braithwaite et al. [37]
	Increased acuity of care	(Unplanned) transfer to a higher level of care (to intensive, semi-intensive, special, intermediate, or acute care unit)	Mills [8], Hiatt et al. [10], Brennan et al. [11], Wilson et al. [13], Baker et al. [16], Classen et al. [18] (GTT), De Meester et al. [21], Hwang et al. [22], O'Leary et al. [23], Pavão et al. [24], Wilson et al. [25], Letaief et al. [26], Cappuccio et al. [28], Soop et al. [29], Kobayashi et al. [30], Mitchell et al. [31], Sari et al. [33], Zegers et al. [34], Resar et al. [35], Herrera-Kiengelher et al. [36], Michel et al. [39], Murff et al. [41], Wolff et al. [42], Thomas et al. [43], Bates et al. [44], Wolff [45], Bates et al. [46], Craddick and Bader [48]
	Intubation	Intubation/re-intubation	Classen et al. [18], Hwang et al. [22], Naessens et al. [27], Resar et al. [35]
	New dialysis	New dialysis	O'Leary et al. [23], Resar et al. [35]
	Unexpected surgery or other procedure	Unplanned visit to operating room or elsewhere for procedure	Williams et al. [32], Forster et al. [38], Michel et al. [39], Chapman et al. [40]
	Change of code status	Code status change in the unit	Resar et al. [35]
Unexpected time course of illness: Ongoing symptoms necessitating re-presentation	Subsequent readmission	(Unplanned) readmission after discharge	Mills [8], Hiatt et al. [10], Brennan et al. [11], Wilson et al. [13], Baker et al. [16], O'Leary et al. [23], Wilson et al. [25], Letaief et al. [26], Soop et al. [29], Zegers et al. [34], Thomas et al. [43], Bates et al. [44]
		Readmission within 12 months	Pavão et al. [24], Kobayashi et al. [30]
		Readmission within 30 days	Classen et al. [18], Hwang et al. [22], Naessens et al. [27], Cappuccio et al. [28]
		Readmission within 28 days	Wolff [45]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Criteria amenable to automated detection			
Category	**Unique concepts**	**Specific wording**	**Study(ies) used**
		Readmission within 21 days	Wolff et al. [42]
		Readmission within 15 days	Herrera-Kiengelher et al. [36]
		Readmission within 72 h	Mitchell et al. [31]
	Readmission causally associated with first admission	Unplanned readmission related to the care provided in the index admission	Sari et al. [33]
	Subsequent outpatient or ED visit because of complications	Subsequent visit to ER or outpatient doctor for complication or adverse results related to this hospitalization	Craddick and Bader [48]
Unexpected time course of illness: Prolonged hospitalization	LOS longer than threshold number of days	Length of stay >35 days	Wolff [45]
		LOS >30 days	Herrera-Kiengelher et al. [36]
		LOS >21 days	Wolff et al. [42]
		LOS >10 days	Williams et al. [32], Chapman et al. [40]
		Unspecified threshold, to be determined by hospital	Craddick and Bader [48]
		>7 days in ICU	Resar et al. [35]
	LOS longer than expected	LOS longer than expected	Williams et al. [32], Chapman et al. [40]
		LOS >50% longer than expected	Cihangir et al. [20]
	LOS longer than percentile for DRG	LOS >90th percentile for DRG in patients under 70, and 95th percentile in those 70 or older	Hiatt et al. [10]
		LOS >90th percentile	Mills [8]
		Unspecified threshold, to be determined by hospital	Craddick and Bader [48]
Change of management: Facility or provider	Change of facility	Unplanned transfer to another hospital/acute care hospital or facility	Mills [8], Hiatt et al. [10], Brennan et al. [11], Wilson et al. [13], Baker et al. [16], O'Leary et al. [23], Pavão et al. [24], Wilson et al. [25], Letaief et al. [26], Soop et al. [29], Kobayashi et al. [30], Williams et al. [32], Sari et al. [33], Herrera-Kiengelher et al. [36], Michel et al. [39], Chapman et al. [40], Wolff et al. [42], Thomas et al. [43], Bates et al. [44], Wolff [45], Wolff [45], Craddick and Bader [48]
	Change of physician or team	Abrupt change of physician in charge	Resar et al. [35]
Change of management: Specific treatment plan	Change of medical treatment	Abrupt medication stop	Classen et al. [18], Hwang et al. [22], O'Leary et al. [23], Naessens et al. [27], Cappuccio et al. [28]
Change of management: Change in procedure diagnosis or type	Change of procedural treatment	Change in procedure	Classen et al. [18], Hwang et al. [22], Naessens et al. [27]
	Cancellation of procedure	Patient booked for surgery and cancelled	Wolff [45]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Criteria amenable to automated detection			
Category	**Unique concepts**	**Specific wording**	**Study(ies) used**
	Difference between diagnosis and pathology results	Pathology report normal or unrelated to diagnosis	Classen et al. [18], Naessens et al. [27]
Diagnostic uncertainty	Multiple consultations	3 or more consultants	Resar et al. [35]
Criteria not likely available in electronic format			
Clinical judgment needed for interpretation	Delays in diagnosis/detection	Diagnosis significantly delayed at any stage of admission	Williams et al. [32]
	Other diagnostic error	Diagnostic error – missed, delayed, misdiagnosis	Mitchell et al. [31]
	Delay in initiating effective treatment	Significant delay in diagnosis/initiating effective treatment at any stage of admission	Chapman et al. [40]
	Abnormal results not addressed by provider	Abnormal laboratory, medical imaging, physical findings or other tests not followed up or addressed	Mitchell et al. [31], Craddick and Bader [48]
	Inadequate observation	Patient deterioration, death, or medical emergency team referral after inadequate observation process	Mitchell et al. [31]
	Undefined deterioration	Worsening condition	Herrera-Kiengelher et al. [36]
		Documented pain or psychological or social injury	Michel et al. [39]
Dissatisfaction with care	Patient dissatisfaction	Dissatisfaction with care received as documented on patient record, or evidence of complaint lodged	Wilson et al. [13], Baker et al. [16], Pavão et al. [24], Letaief et al. [26], Soop et al. [29], Kobayashi et al. [30], Williams et al. [32], Sari et al. [33], Zegers et al. [34], Herrera-Kiengelher et al. [36], Michel et al. [39], Chapman et al. [40]
	Family complaints about care	Relative made complaint regarding care	Wilson et al. [25], Williams et al. [32], Resar et al. [35], Forster et al. [38], Michel et al. [39], Chapman et al. [40], Craddick and Bader [48]
	Patient pursuing litigation	Documentation or correspondence suggesting/ indicating litigation (either contemplated or actual)	Hiatt et al. [10], Wilson et al. [13], Baker et al. [16], Pavão et al. [24], Wilson et al. [25], Letaief et al. [26], Soop et al. [29], Kobayashi et al. [30], Sari et al. [33], Zegers et al. [34], Herrera-Kiengelher et al. [36], Forster et al. [38], Michel et al. [39], Thomas et al. [43], Bates et al. [44]
	Provider dissatisfaction	Doctor or nurse unhappy about any aspect of care	Williams et al. [32], Chapman et al. [40]
Reporting external to EHR	Referral to hospil ethics board	Specific case referral	Mitchell et al. [31]
	Incident report to external board	High-level incident report	Mitchell et al. [31]