Electronic Health Record-Based Surveillance of Diagnostic Errors in Primary Care

Hardeep Singh, MD, MPH[1], **Traber Davis Giardina, MA, MSW**[1], **Samuel N. Forjuoh, MD, MPH, DrPH**[2], **Michael D. Reis, MD**[2], **Steven Kosmach, MSN, RN**[3], **Myrna M. Khan, PhD, MBA**[1], and **Eric J. Thomas, MD, MPH**[4]

[1]Houston VA HSR&D Center of Excellence, Michael E. DeBakey Veterans Affairs Medical Center and the Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, Texas

[2]Department of Family & Community Medicine, Scott & White Healthcare, Texas A&M Health Science Center, College of Medicine, Temple, TX

[3]Division of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA

[4]University of Texas at Houston – Memorial Hermann Center for Healthcare Quality and Safety, and Division of General Medicine, Department of Medicine, University of Texas Medical School at Houston, Houston, Texas

Abstract

Background—Diagnostic errors in primary care are harmful but difficult to detect. We tested an electronic health record (EHR)-based method to detect diagnostic errors in routine primary care practice.

Methods—We conducted a retrospective study of primary care visit records "triggered" through electronic queries for possible evidence of diagnostic errors: Trigger 1: A primary care index visit followed by unplanned hospitalization within 14 days; and Trigger 2: A primary care index visit followed by   1 unscheduled visit(s) within 14 days. Control visits met neither criterion. Electronic trigger queries were applied to EHR repositories at two large healthcare systems between October 1, 2006 and September 30, 2007. Blinded physician-reviewers independently determined presence or absence of diagnostic errors in selected triggered and control visits. An error was defined as a missed opportunity to make or pursue the correct diagnosis when adequate data was available at the index visit. Disagreements were resolved by an independent third reviewer.

Results—Queries were applied to 212,165 visits. On record review, we found diagnostic errors in 141 of 674 Trigger 1-positive records (PPV=20.9%, 95% CI, 17.9%-24.0%) and 36 of 669

Address for Correspondence and Reprints: Hardeep Singh, MD, MPH, VA Medical Center (152), 2002 Holcombe Blvd, Houston, TX 77030, Phone (713) 794-8515, hardeeps@bcm.edu, Fax: 713-748-7359.

Trigger 2-positive records (PPV=5.4%, 95% CI, 3.7%-7.1%). The control PPV of 2.1% (95% CI, 0.1%-3.3%) was significantly lower than that of both triggers (P .002). Inter-rater reliability was modest, though higher than in comparable previous studies ($\kappa = 0.37$ [95% CI=0.31-0.44]).

Conclusions—While physician agreement on diagnostic error remains low, an EHR-facilitated surveillance methodology could be useful for gaining insight into the origin of these errors.

Keywords

diagnostic errors; primary care; patient safety; electronic health records; triggers; automated surveillance; error detection

Background

Although certain types of medical errors (such as diagnostic errors) are likely to be prevalent in primary care, medical errors in this setting are generally understudied.[1, 1-7] Data from outpatient malpractice claims [2, 8-10] consistently rank missed, delayed, and wrong diagnoses as the most common identifiable error. Other types of studies have also documented the magnitude and significance of diagnostic errors in outpatient settings.[11-17] Although these data point to an important problem, diagnostic errors have not been studied as well as other types of errors.[18-20] These errors are difficult to detect and define[9] and physicians might not always agree on the occurrence of error. Methods to improve detection and learning from diagnostic errors are key to advancing their understanding and prevention.[19, 21, 22]

Existing methods for diagnostic error detection (e.g., random chart reviews, voluntary reporting, claims file review, etc.) are inefficient, biased, or unreliable.[23] In our preliminary work, we developed two computerized triggers to identify primary care patient records that may contain evidence of trainee-related diagnostic errors.[24] Triggers are signals that can alert providers to review the record to determine whether a patient safety event occurred.[25-29] Our triggers were based on the rationale that an unexpected hospitalization or return clinic visit after an initial primary care visit may indicate that a diagnosis was missed during the first visit. Although the positive predictive value (PPV) was modest (16.1% and 9.7% for the two triggers, respectively), it was comparable to that of previously designed electronic triggers to detect potential ambulatory medication errors.[30, 31] Our previous findings were limited by a lack of generalizability outside of the study setting (a Veterans Affairs [VA] internal medicine trainee clinic), poor agreement between reviewers on presence of diagnostic error, and a high proportion of false positive cases that led to unnecessary record reviews.

In this study, we refined our prior approach and evaluated a methodology to improve systems-based detection of diagnostic errors in routine primary care. Our ultimate goal was to create a surveillance method that primary care practices could adopt in the future in order to start addressing the burden of diagnostic errors.

Methods

Design and Settings

We designed electronic queries (triggers) to detect patterns of visits that could have been precipitated by diagnostic errors and applied these queries to all primary care visits at two large health systems over a 1-year time period. We performed chart reviews on samples of "triggered" visits (i.e., those that met trigger criteria) and control visits (those that did not meet trigger criteria) to determine the presence or absence of diagnostic errors.

Both study sites provided longitudinal care in a relatively closed system and had integrated and well-established EHRs. Each site's electronic data repository contained updated administrative and clinical data extracted from the EHR. At Site A, a large VA facility, 35 full-time primary care providers (PCPs) saw approximately 50,000 unique patients annually in scheduled primary care follow-up visits and "drop-in" unscheduled or urgent care visits. PCPs included 25 staff physician-internists, about half of whom supervised residents assigned to them two half-days a week; the remaining PCPs were allied health providers. Emergency room (ER) staff provided care after hours and on weekends. At Site B, a large private integrated health care system, 34 PCPs (family medicine physicians) provided care to nearly 50,000 unique patients in 4 community-based clinics, and about two-third supervised family practice residents part-time. Clinic patients sought care at the ER of the parent hospital after-hours. To minimize after-hours attrition to hospitals outside our study settings, we did not apply the trigger to patients assigned to remote satellite clinics of the parent facilities. Both settings included ethnically and socioeconomically diverse patients from rural and urban areas. Local IRB approval was obtained at both sites.

Trigger Application

Using a Structured Query Language (SQL)-based program, we applied three queries to electronic repositories at the two sites to identify primary care index visits (defined as scheduled or unscheduled visits to physicians, physician-trainees, and allied health professionals that did not lead to immediate hospitalizations) between October 1, 2006 and September 30, 2007 that met the following criteria:

Trigger 1: A primary care visit followed by an unplanned hospitalization that occurred between 24 hours and 14 days after the visit.

Trigger 2: A primary care visit followed by 1 or more unscheduled primary care visits, an urgent care visit, or an ER visit that occurred within 14 days (excluding Trigger 1-positive index visits).

Controls: All primary care visits from the study period that did not meet either trigger criterion.

The triggers above were based on our previous work and refined to improve their performance (Table 1). Our pilot reviews suggested that when a 3 or 4 week interval between index and return visits was used, reasons for return visits were less clearly linked to index visit and less attributable to error. Thus, a 14-day cut-off was chosen. In addition, we attempted to electronically remove records with false positive index visits, such as those associated with planned hospitalizations.

Data Collection and Error Assessment

We performed detailed chart reviews on selected triggered and control visits. If patients met a trigger criterion more than once, only one (earliest) index visit was included (unique patient record). Some records did not meet criteria for detailed review because the probability of us being able to identify an error at the index visit using this methodology would be nil for one of the following reasons: absence of any clinical documentation at index visit; patients left clinic without being seen at the index visit; patient only saw a nurse, dietician or social worker; or patients were advised hospitalization at index visit for further evaluation but refused. For the purposes of our analysis, these records were categorized as false positives, even though some of these could contain diagnostic errors.

Eligible unique records were randomly assigned to trained physician-reviewers from outside institutions. Reviewers were chief residents and clinical fellows (medicine subspecialties) and were selected based on faculty recommendations and interviews by the research team.

They were blinded to the goals of the study and to the presence or absence of triggers. All reviewers underwent stringent quality control and pilot testing procedures and reviewed 25-30 records each before they started collecting study data. Through several sessions, we trained the reviewers to determine errors at the index visit through a detailed review of the EHR about care processes involving the index visit and subsequent visits. Reviewers were also instructed to review EHR data from subsequent months after the index visit to help verify whether a diagnostic error was made. Although we used an explicit definition of diagnostic error from the literature [32] and a structured training process based upon our previous record review studies, assessment of the diagnostic process involves implicit judgments. To improve reliability and to operationalize the definition of diagnostic error, we asked reviewers to judge diagnostic performance based strictly on data either already available or easily available to the treating provider at the time of the index clinic visit (i.e., adequate "specific" data must have been available at the time to either make or pursue the correct diagnosis). Thus, reviewers were asked to judge errors when missed opportunities to make an earlier diagnosis occurred. A typical example of a missed opportunity is when adequate data to suggest the final, correct diagnosis was already present at the index visit (eg, constellation of certain findings such as dyspnea, elevated venous pressures, pedal edema, chest crackles, or pulmonary edema on chest-x ray should suggest heart failure). Another common example of a missed opportunity is when certain documented abnormal findings (eg, cough, fever, and dyspnea) should have prompted additional evaluation (eg. chest x-ray).

If the correct diagnosis was documented, and the patient was advised outpatient treatment (versus hospitalization) but returned within 14 days and got hospitalized anyway, reviewers did not attribute such provider management decisions to diagnostic error. Each record was initially reviewed by two independent reviewers. Because we expected a number of ambiguous situations, when two reviewers disagreed on presence of diagnostic error, a third, blinded review was used to make the final determination.[33] Charts were randomly assigned to available reviewers, about 50 charts at a time. Not all reviewers were always available due to clinical and personal commitments.

A structured data collection form, adapted from our previous work,[24] was used to record documented signs and symptoms, clinician assessment and diagnostic decisions. Both sites have well-structured procedures to scan reports and records from physicians external to the system into the EHR and thus information about patient care outside our study settings was also reviewed when available. To reduce hindsight bias,[34, 35] we did not ask reviewers to assess patient harm. We computed Cohen's kappa (κ) to assess inter-reviewer agreement prior to tiebreaker decisions.

Sampling Strategy

To determine our sample size, we focused exclusively on determining the number of Trigger 1 records because of its higher PPV and potential for wider use. We initially calculated the sample size based on our anticipation that by refining trigger criteria we could lower the proportion of false positive cases for Trigger 1 to 20%, compared to the false positive proportion of 34.1% in our previous work.[24] We estimated a minimum sample size of 310 to demonstrate a significant difference ($P < .05$) in the false positive proportion between the new and previous Trigger 1 with 80% power. We further refined the sample size in order to detect an adequate number of errors to allow future sub-classification of error type and contributory factors, consistent with the sample size of 100 error cases used in a landmark study on diagnostic error.[32] Anticipating that we would still be able to obtain a PPV of at least 16.1% for Trigger 1 (as in our previous study), we estimated that at least 630 patient visits meeting Trigger 1 criteria would be needed to discover 100 error cases. However, on initial test queries of the data repository at Site B, we found only 220 unique records positive

for Trigger 1 in the study period, whereas at Site A it was much higher. We thus used all 220 Trigger 1-positive records from Site B and randomly selected the remaining charts from Site A to achieve an adequate sample size for Trigger 1, oversampling by about 10% to allow for any unusable records.[24] We randomly selected comparable numbers of Trigger 2-positive records but selected slightly fewer unique records for controls because we expected fewer manual exclusions related to situations when patients were advised hospitalization but refused and elected to return a few days later (trigger false positives). . For both Trigger 2-positive records and controls, we maintained the sampling ratio of Trigger 1; thus about two-thirds of the records were from Site A.

Data Analysis

We calculated PPVs for both triggers and compared these with PPVs for controls. We also calculated the proportion of false positive cases for each trigger and compared them to our previously used methods. We tallied the frequency of clinical conditions associated with the diagnostic errors that we discovered. The z test was used to test the equality of proportions (for PPV or false positives) when comparing between sites and prior study results. When lower confidence intervals (CIs) were negative due to small sample size, we calculated exact CIs.

Results

We applied the triggers to 212,165 primary care visits (106,143 at Site A and 106,022 at Site B) that contained 81,483 unique patient records. Our sampling strategy resulted in 674 positive unique patient records for Trigger 1, 669 positive unique patient records for Trigger 2, and 614 unique control patient records for review (Figure 1). On detailed review, diagnostic errors were found in 141 Trigger 1 positive records and 36 Trigger 2 positive records, yielding a PPV of 20.9% for Trigger 1 (95% CI, 17.9%-24.0%) and 5.4% for Trigger 2 (95% CI, 3.7%-7.1%). Errors were found in 13 control records. The control PPV of 2.1% (95% CI, 0.1%-3.3%) was significantly lower than that of both Trigger 1 ($P < .001$) and Trigger 2 ($P = .002$). Trigger PPVs were equivalent between sites ($P = .9$ for both triggers).

Prior to the tiebreaker, kappa agreement in triggered charts was 0.37 (95% CI=0.31-0.44). Of 285 charts with disagreement, the third reviewer detected a diagnostic error in 29.8%. In 96% of triggered error cases, the reviewers established a relationship between the admission or second outpatient visit and the index visit. Figure 2 shows the distribution of diagnostic errors in the inclusion sample over time interval between index and return dates for both Trigger 1 and Trigger 2 records.

The proportion of false positive cases were not statistically different between two sites (Table 2). The overall proportion of false positives was 15.6% for Trigger 1 and 9.6% for Trigger 2, significantly lower than those in our previous study (34.1% and 25.0%, respectively, $P < .001$ for both comparisons). Because many false positives (no documentation, telephone or non-PCP encounters, etc.) could potentially be coded accurately and identified electronically through information systems, we estimated the highest PPV potentially achievable by an ideal system that screened out those encounters. Our estimates suggest that Trigger 1 PPV would increase from 20.9% (CI 17.9-24.2%) to 24.8% (CI 21.3-28.5%) if electronic exclusion of false positives was possible.

Three types of scenarios occurred as a result of review procedures: 1) both initial reviewers agreed that it was an error, 2) the independent third reviewer judged it to be an error after initial disagreement, and 3) the independent third reviewer judged it not to be an error. The examples in Table 3 illustrate several reasons why reviewers initially disagreed and support

why using more than one reviewer (and as many as three at times) is essential to making diagnostic error assessment more reliable.

Discussion

We evaluated a trigger methodology to rigorously improve surveillance of diagnostic errors in routine primary care practice. One of our computerized triggers had a higher PPV to detect diagnostic error, and lower proportion of false positives, than any other known method. Additionally, the reliability of diagnostic error detection in our triggered population was higher than previous studies on diagnostic error.[36] These methods can be used to identify and learn from diagnostic errors in primary care so that preventive strategies can be developed.

Our study has several unique strengths. We leveraged the electronic health record (EHR) infrastructure of two large healthcare systems that involved several types of practice settings (internal medicine and family medicine, academic and nonacademic). The increasing use of electronic health records (EHRs) facilitates creation of health data repositories that contain longitudinal patient information in a far more integrated fashion than in previous record systems.

Because of the heterogeneous causes and outcomes of diagnostic errors, several types of methods are needed to capture the full range of these events. [2, 20, 37] Our trigger methodology thus could have broad applicability especially because our queries contained information available in almost all EHRs.

The study has key implications for future primary care reform. Given high patient volumes, rushed office visits, and multiple competing demands for PCPs' attention, our findings are not surprising and call for a multi-pronged intervention effort for error prevention.[21, 38, 39] Primary care quality improvement initiatives should consider using active surveillance methods such as Trigger 1, an approach that could be equated to initiatives related to electronic surveillance for medication errors and nosocomial infections.[25, 26, 40, 41] For instance, these techniques can be used for oversight and monitoring of diagnostic performance with feedback to frontline practitioners about errors and missed opportunities. A review of triggered cases by primary care teams to ensure that all contributing factors are identified – not just those related to provider performance – will foster interdisciplinary quality improvement. This strategy could complement and strengthen other provider-focused interventions, which in isolation are unlikely to effect significant change.

Underdeveloped detection methods have been a major impediment to progress in understanding and preventing diagnostic errors. By refining our triggers and reducing false positives, we created detection methods that are far more efficient than conducting random record reviews or relying on incident reporting systems.[42] Previously used methods to study diagnostic errors have notable limitations: autopsies are now infrequent,[43] self-report methods (eg, surveys and voluntary reporting) are prone to reporting bias, and malpractice claims, although useful, shed light on a narrow and non-representative spectrum of medical error.[2, 15, 20, 23, 44] Medical record review is often considered the gold standard for studying diagnostic errors, but it is time consuming and costly. Moreover, random record review has a relatively low yield if used non-selectively, as shown in our non-triggered (control) group.[45] While our methodology can be useful to "trigger" additional investigation, there are challenges to reliable diagnostic error assessment such as low agreement rates and uncertainty about how best to statistically evaluate agreement in the case of low error rates.[46, 47] Thus, although our methods improve detection of diagnostic errors, their ultimate usefulness will depend on continued efforts to improve their reliability.

Our findings have several limitations. Our methods may not be generalizable to primary care practices that do not belong to an integrated health system or that lack staffing necessary for record reviews. Although chart review may be the best available method for detecting diagnostic errors, it is not perfect because written documentation might not accurately represent all aspects of a patient encounter. The kappa agreement between our reviewers was only fair. However, judgment for diagnostic errors is more difficult than other types of errors,[20] and our kappa was higher than in other comparable studies of diagnostic error.[48] This methodology might not be able to detect many types of serious diagnostic errors that would not result in another medical encounter within 14 days..[49] We also likely underestimated the number of errors because we were unable to access admissions or outpatient visits at other institutions, and because some misdiagnosed patients, unknown to us, might have recovered without further care (i.e. false negative situations). Finally, we did not assess severity of errors or associated harm. However, the fact that these errors led to further contact with the healthcare system suggests they were not inconsequential and would have been defined as adverse events in most studies.

In summary, an EHR-facilitated trigger and review methodology can be used for improving detection of diagnostic errors in routine primary care practice. Primary care reform initiatives should consider these methods for error surveillance, a key first step toward error prevention.

Acknowledgments

References

1. Gandhi TK, Lee TH. Patient Safety beyond the Hospital. New England Journal of Medicine. 2010 Sep 8; 363(11):1001–3. [PubMed: 20825311]

2. Phillips R Jr, Bartholomew LA, Dovey SM, Fryer GE Jr, Miyoshi TJ, Green LA. Learning from malpractice claims about negligent, adverse events in primary care in the United States. Qual Saf Health Care. 2004 Apr; 13(2):121–6. [PubMed: 15069219]

3. Bhasale A. The wrong diagnosis: identifying causes of potentially adverse events in general practice using incident monitoring. Fam Pract. 1998 Aug; 15(4):308–18. [PubMed: 9792345]

4. Elder NC, Dovey SM. Classification of medical errors and preventable adverse events in primary care: a synthesis of the literature. J Fam Pract. 2002 Nov; 51(11):927–32. [PubMed: 12485545]

5. Woods DM, Thomas EJ, Holl JL, Weiss KB, Brennan TA. Ambulatory care adverse events and preventable adverse events leading to a hospital admission. Quality and Safety in Healthcare. 2007

6. Wachter RM. Is ambulatory patient safety just like hospital safety, only without the "stat"? Ann Intern Med. 2006 Oct 3; 145(7):547–9. [PubMed: 17015874]

7. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998 Jul 20; 169(2):73–6. [PubMed: 9700340]

8. Chandra A, Nundy S, Seabury SA. The growth of physician medical malpractice payments: evidence from the National Practitioner Data Bank. Health Aff (Millwood). 2005 Jan.:W5-240–W5-249. Suppl Web Exclusives. [PubMed: 15928255]

9. Graber M. Diagnostic errors in medicine: a case of neglect. Jt Comm J Qual Patient Saf. 2005 Feb; 31(2):106–13. [PubMed: 15791770]

10. Bishop TF, Ryan AK, Casalino LP. Paid malpractice claims for adverse events in inpatient and outpatient settings. JAMA. 2011 Jun 15; 305(23):2427–31. [PubMed: 21673294]

11. Casalino LP, Dunham D, Chin MH, et al. Frequency of Failure to Inform Patients of Clinically Significant Outpatient Test Results. Arch Intern Med. 2009 Jun 15; 169(12):1123–9. [PubMed: 19546413]

12. Schiff GD, Hasan O, Kim S, et al. Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors. Arch Intern Med. 2009 Nov 9; 169(20):1881–7. [PubMed: 19901140]

13. Singh H, Thomas EJ, Mani S, et al. Timely follow-up of abnormal diagnostic imaging test results in an outpatient setting: are electronic medical records achieving their potential? Arch Intern Med. 2009 Sep 28; 169(17):1578–86. [PubMed: 19786677]

14. Singh H, Hirani K, Kadiyala H, et al. Characteristics and predictors of missed opportunities in lung cancer diagnosis: an electronic health record-based study. J Clin Oncol. 2010 Jul 15; 28(20):3307–15. [PubMed: 20530272]

15. Singh H, Thomas EJ, Wilson L, et al. Errors of diagnosis in pediatric practice: a multisite survey. Pediatrics. 2010 Jul; 126(1):70–9. [PubMed: 20566604]

16. Singh H, Thomas EJ, Sittig DF, et al. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? Am J Med. 2010 Mar; 123(3):238–44. [PubMed: 20193832]

17. Singh H, Daci K, Petersen L, et al. Missed opportunities to initiate endoscopic evaluation for colorectal cancer diagnosis. Am J Gastroenterol. 2009; 104(10):2543–54. [PubMed: 19550418]

18. Schiff GD, Kim S, Abrams R, Cosby K, Lambert B, Elstein AS. Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project. Advances in Patient Safety. 2005 Feb 1.

19. Wachter RM. Why diagnostic errors don't get any respect--and what can be done about them. Health Aff (Millwood). 2010 Sep; 29(9):1605–10. [PubMed: 20820015]

20. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. Ann Intern Med. 2006 Oct 3; 145(7):488–96. [PubMed: 17015866]

21. Singh H, Graber M. Reducing Diagnostic Error through Medical Home-Based Primary Care Reform. JAMA. 2010

22. Newman-Toker DE, Pronovost PJ. Diagnostic Errors--The Next Frontier for Patient Safety. JAMA. 2009 Mar 11; 301(10):1060–2. [PubMed: 19278949]

23. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. J Gen Intern Med. 2003 Jan; 18(1):61–7. [PubMed: 12534766]

24. Singh H, Thomas E, Khan MM, Petersen L. Identifying diagnostic errors in primary care using an electronic screening algorithm. Arch Intern Med. 2007 Feb 7; 167(3):302–8. [PubMed: 17296888]

25. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. JAMA. 1991 Nov 27; 266(20):2847–51. [PubMed: 1942452]

26. Szekendi MK, Sullivan C, Bobb A, et al. Active surveillance using electronic triggers to detect adverse events in hospitalized patients. Qual Saf Health Care. 2006 Jun; 15(3):184–90. [PubMed: 16751468]

27. Kaafarani HMA, Rosen AK, Nebeker JR, et al. Development of trigger tools for surveillance of adverse events in ambulatory surgery. Qual Saf Health Care.

28. Handler SM, Hanlon JT. Detecting Adverse Drug Events Using a Nursing Home Specific Trigger Tool. Ann Longterm Care. 2010 May; 18(5):17–22. [PubMed: 20808714]

29. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood). 2011 Apr; 30(4):581–9. [PubMed: 21471476]

30. Field TS, Gurwitz JH, Harrold LR, et al. Strategies for detecting adverse drug events among older persons in the ambulatory setting. J Am Med Inform Assoc. 2004 Nov; 11(6):492–8. [PubMed: 15299000]

31. Honigman B, Lee J, Rothschild J, et al. Using computerized data to identify adverse drug events in outpatients. J Am Med Inform Assoc. 2001 May; 8(3):254–66. [PubMed: 11320070]

32. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005 Jul 15; 165(13):1493–9. [PubMed: 16009864]

33. Forster AJ, O'Rourke K, Shojania KG, van Walraven C. Combining ratings from multiple physician reviewers helped to overcome the uncertainty associated with adverse event classification. Journal of Clinical Epidemiology. 2007 Sep; 60(9):892–901. [PubMed: 17689805]

34. Fischhoff B. Hindsight not equal to foresight: the effect of outcome knowledge on judgment under uncertainty 1975. Qual Saf Health Care. 2003 Aug 1; 12(4):304–11. [PubMed: 12897366]

35. McNutt R, Abrams R, Hasler S. Diagnosing diagnostic mistakes: AHRQ web morbidity and mortality rounds. AHRQ. 2005 9-25-2006. Ref Type: Internet Communication.

36. Zwaan L, de Bruijne M, Wagner C, et al. Patient Record Review of the Incidence, Consequences, and Causes of Diagnostic Adverse Events. Arch Intern Med. 2010 Jun 15; 170(12):1015–21. [PubMed: 20585065]

37. Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: current understanding and future directions. J Clin Oncol. 2007 Nov 1; 25(31):5009–18. [PubMed: 17971601]

38. Singh H, Weingart S. Diagnostic Errors in Ambulatory Care: Dimensions and Preventive Strategies. Adv Health Sci Educ Theory Pract. 2009; 14(1):57–61. [PubMed: 19669923]

39. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. Qual Saf Health Care. 2006 Jun 15; 15(3):159–64. [PubMed: 16751463]

40. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. J Am Med Inform Assoc. 2003 Mar; 10(2):115–28. [PubMed: 12595401]

41. Jha AK, Kuperman GJ, Teich JM, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. J Am Med Inform Assoc. 1998 May; 5(3):305–14. [PubMed: 9609500]

42. Sevdalis N, Jacklin R, Arora S, Vincent CA, Thomson RG. Diagnostic error in a national incident reporting system in the UK. J Eval Clin Pract. 2010 Aug 19.

43. Shojania KG, Burton EC. The Vanishing Nonforensic Autopsy. New England Journal of Medicine. 2009 Jun 15; 358(9):873–5. [PubMed: 18305264]

44. McAbee GN, Donn SM, Mendelson RA, McDonnell WM, Gonzalez JL, Ake JK. Medical Diagnoses Commonly Associated With Pediatric Malpractice Lawsuits in the United States. Pediatrics. 2008 Dec 1.122(6):e1282, e1286. [PubMed: 19047227]

45. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado in 1992. Medical Care. 2000; 38:261–71. [PubMed: 10718351]

46. Localio AR, Weaver SL, Landis JR, et al. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. Ann Intern Med. 1996 Sep 15; 125(6):457–64. [PubMed: 8779457]

47. Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Ann Intern Med. 2002 Jun 15; 136(11):812–6. [PubMed: 12044129]

48. Zwaan L, de Bruijne M, Wagner C, et al. Patient Record Review of the Incidence, Consequences, and Causes of Diagnostic Adverse Events. Arch Intern Med. 2010 Jun 15; 170(12):1015–21. [PubMed: 20585065]

49. Shojania K. The Elephant of Patient Safety: What You See Depends on How You Look. Joint Commission Journal on Quality and Patient Safety. 2010; 36(9):399–401. [PubMed: 20873672]
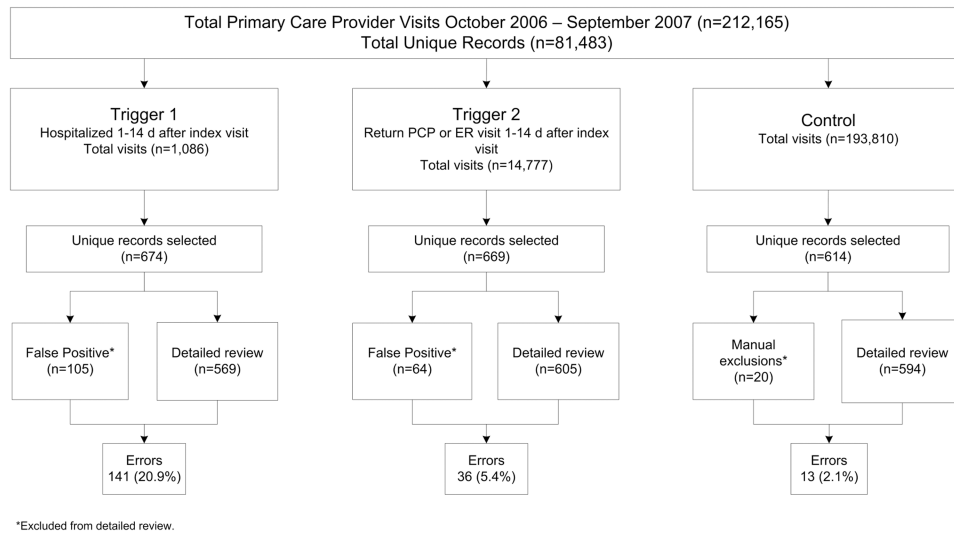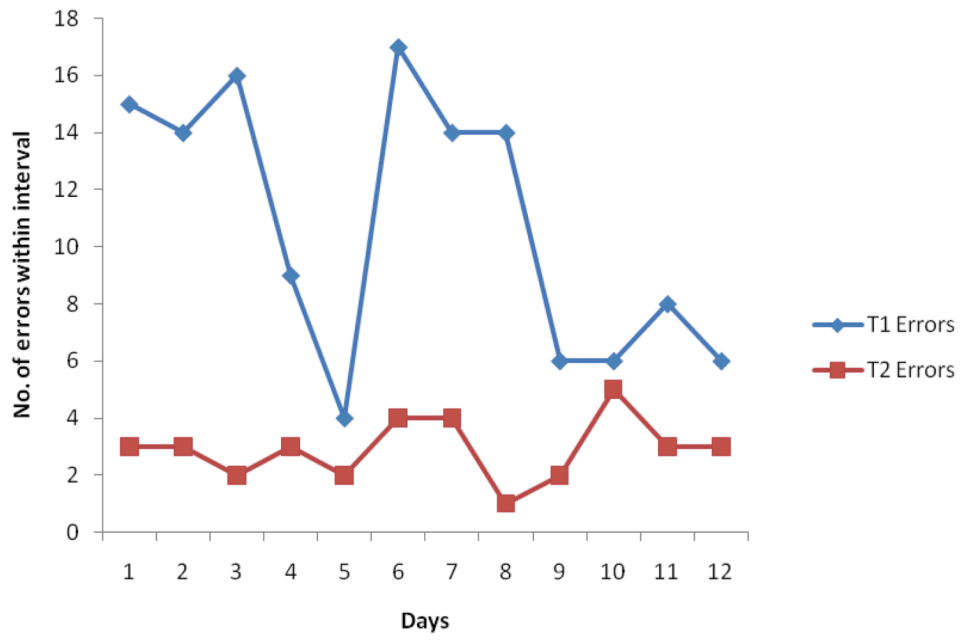
Figure 1. Study flowchart

Figure 2. Number of Errors per Day post-Index Visit in the Triggered Subset

Table 1

Rationale of Trigger Modifications from Previous Work[24]

Trigger Characteristics	Previous trigger	New trigger	Rationale
Time period	10 days	14 days	Previous findings showed that diagnostic errors continued to be discovered at the 10 day cut-off.
	Did not account for planned hospitalizations or elective surgeries	Electronically excludes planned or elective admissions related to (or from) day surgery, scheduled ambulatory admit, pre-operative clinics, cardiology invasive procedure clinic	
Inclusion criteria	Did not account for admissions to units considered "non-acute"	Electronically excludes admissions to long term and intermediate care, and rehabilitation units	Lower proportion of triggered false positives will increase PPV as well as efficiency of record reviews.
	Included all hospitalizations	Includes only hospitalizations electronically designated as "acute care" (e.g. acute medicine, acute surgery, acute mental health)	

Table 2

Site Specific Positive Predictive Values (PPVs) and False Positive Proportions

	Total number of errors	Total number selected for review	PPV in current EHR system	95% Exact CIs	False positive proportion[¥]		95% Exact CIs	Expected PPV in an ideal EHR system[*]	95% Exact CIs
	n	n	%		n	%		%	
Site A T1	95	454	20.9%	(17.3, 25.0)	74	16.3%	(13.0, 20.0)	25.0%	(20.7, 29.7)
Site A T2	23	432	5.3%	(3.4, 7.9)	50	11.6%	(8.7, 15.0)	6.0%	(3.9, 8.9)
Site A Control	11	414	2.7%	(1.3, 4.7)	7	1.7%	(0.70, 3.4)	2.7%	(1.4, 4.8)
Site B T1	46	220	20.9%	(15.7, 26.9)	31	14.1%	(9.8, 19.4)	24.3%	(18.4, 31.1)
Site B T2	13	237	5.5%	(3.0, 9.2)	14	5.9%	(3.3, 9.7)	5.8%	(3.1, 9.8)
Site B Control	2	200	1.0%	(0.12, 3.6)	13	6.5%	(3.5, 10.9)	1.1%	(0.13, 3.8)
Overall	190	1957			189				

[*] Calculation: Numerator is the total number of errors and Denominator is the total number of selected for review minus the number of false positives

[¥] Includes records where no documentation of clinical notes was found, patients left without being seen by the provider, patients saw a non-provider, or patients were advised admission for further work-up but refused

Abbreviations: EHR, electronic health record; T1, Trigger 1; T2, Trigger 2.

Table 3

Brief Vignettes to Illustrate Three Types of Reviewer Agreement

No.	Vignette	Provider Diagnosis	Reviewer 1	Reviewer 2	Reviewer 3
1.	88 y/o male with h/o lymphoma presented with headache, cough, green sputum, and fever. Provider did not order labs or x-ray to evaluate for pneumonia.	Acute Bronchitis	Missed pneumonia	Missed pneumonia	N/A
2.	61 y/o male with h/o hypertension and neck pain presented with intermittent numbness, weakness, and tingling of both hands	Peripheral neuropathy	Missed cervical myelopathy with cord compression	Missed cervical cord compression	N/A
3.	45 y/o female with cough and fever; diagnosed with pharyngolaryngo-tracheitis. PCP read chest x-ray as normal; patient returned with continued symptoms and found to have lobar pneumonia on initial x-ray that had been read by the radiologist.	Tracheitis	Pneumonia	No error	Pneumonia
4.	87 y/o male with right hand swelling, pain and new onset bilateral decreased grip; diagnosed few weeks later with carpel tunnel syndrome.	Osteoarthritis	No error	Missed carpal tunnel syndrome	Missed carpal tunnel syndrome
5.	65 y/o male with left hand swelling and erythema after a small cut; treated for cellulitis at index visit. Returned 4 days later with "failure to respond" and admitted for IV antibiotics with some improvement. Uric acid found elevated and thus additionally treated for gout; given prednisone.	Cellulitis	No error	Missed gout	No error
6.	42 y/o female with post hospitalization follow-up for CHF exacerbation c/o 1 week shortness of breath and congestion. Treated for URI but Lasix increased to address CHF. Symptoms progressed so patient admitted for CHF exacerbation/URI.	CHF and bronchitis	Missed CHF exacerbation	No error	No error

Note. 1-2 are case scenarios where both initial reviewers agreed on error; 3-4 are case scenarios when the independent third reviewer judged it to be an error after initial disagreement, and 5-6 are case scenarios when the independent third reviewer judged it not to be an error (after initial disagreement).

Abbreviations: c/o, complaint of; CHF, chronic heart failure; h/o, history of; NA, not available; PCP, primary care provider; URI, upper respiratory infection; y/o, 3 year old.