Automated detection of harm in healthcare with information technology: a systematic review

Malavika Govindan,[1] Aricca D Van Citters,[2] Eugene C Nelson,[1] Jane Kelly-Cummings,[3] Gautham Suresh[4]

► An additional appendix is published online only. To view these files please visit the journal online (http://qshc.bmj.com).

[1]The Dartmouth Institute for Health Policy and Clinical Practice, Center for Leadership and Improvement, Dartmouth Medical School, Lebanon, New Hampshire, USA
[2]Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire, USA
[3]Society for Hospital Medicine, Philadelphia, Pennsylvania, USA
[4]Department of Pediatrics, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA

Correspondence to
Dr Gautham Suresh, Dartmouth-Hitchcock Medical Center, One Medical Center Drive, Lebanon, NH 03756, USA;
gautham.suresh@hitchcock.org

Accepted 14 July 2009
Published Online First 29 July 2010

ABSTRACT

Context To improve patient safety, healthcare facilities are focussing on reducing patient harm. Automated harm-detection methods using information technology show promise for efficiently measuring harm. However, there have been few systematic reviews of their effectiveness.

Objective To perform a systematic literature review to identify, describe and evaluate effectiveness of automated inpatient harm-detection methods.

Methods Data sources included MEDLINE and CINAHL databases indexed through August 2008, extended by bibliographic review and search of citing articles. The authors included articles reporting effectiveness of automated inpatient harm-detection methods, as compared with other detection methods. Two independent reviewers used a standardised abstraction sheet to extract data about automated and comparison harm-detection methods, patient samples and events identified. Differences were resolved by discussion.

Results From 176 articles, 43 articles met inclusion criteria: 39 describing field-defined methods, two using natural language processing and two using both methods. Twenty-one studies used automated methods to detect adverse drug events, 10 detected general adverse events, eight detected nosocomial infections, and four detected other specific adverse events. Compared with gold standard chart review, sensitivity and specificity of automated harm-detection methods ranged from 0.10 to 0.94 and 0.23 to 0.98, respectively. Studies used heterogeneous methods that often were flawed.

Conclusion Automated methods of harm detection are feasible and some can potentially detect patient harm efficiently. However, effectiveness varied widely, and most studies had methodological weaknesses. More work is needed to develop and assess these tools before they can yield accurate estimates of harm that can be reliably interpreted and compared.

INTRODUCTION

It is widely recognised that harm caused by the healthcare system is a major source of morbidity and mortality in hospitalised patients.[1] An estimated 15 million instances of medical harm occur in the USA every year.[2] However, the lack of simple, practical and accurate methods to identify adverse events in hospitals has hampered efforts to develop routine monitoring systems, assess the impact of interventions to prevent harm and compare interhospital performance.

Detecting incidence and types of patient harm are prerequisites for implementing strategies to prevent harm. Manual, comprehensive chart review by trained professionals has been used in key studies and can be considered the gold-standard harm-detection method.[3–6] However, this approach requires time and trained abstractors, thereby decreasing its feasibility as a pragmatic method for routine measurement of adverse events.

Several organisations are currently using the Institute for Healthcare Improvement's Global Trigger Tool, which is based on manual chart review, and allows targeted chart review to identify harm more efficiently than comprehensive chart review and more extensively than voluntary reporting of harm.

Automated strategies of harm detection that use computerised methods to scan patient records may require fewer time and personnel resources than traditional methods, and can potentially provide real-time surveillance alerts. We performed this review to: (1) identify types of automated methods of inpatient harm detection described in published literature, (2) describe types of events identified by these methods and (3) evaluate accuracy of these methods in identifying harm. We also independently evaluated the quality and validity of key studies.

METHODS

Definitions

In this review, we used the terms harm, automated harm detection and gold standard chart review as defined in Box 1.

Data sources/study selection

We (MG and AVC) identified articles for this review through a literature search of MEDLINE (start date 1950) and CINAHL (start date 1982) using the following search terms: (harm OR adverse event OR adverse drug event OR nosocomial infection) AND (automated OR computerised OR electronic) AND (identify OR detect OR detection OR recognise OR recognition). We identified additional articles using bibliographic review of key articles, the 'related articles' feature of Medline, and the 'find similar' and 'find citing articles' feature of CINAHL. We reviewed the title and abstract of each article, and obtained the full text of relevant articles. We limited our search to English language articles indexed through 31 August 2008.

We included studies that: (a) occurred in an inpatient setting, (b) described an automated harm-detection method, (c) measured actual harm and (d) compared the automated method to an alternative method of harm detection.

Box 1 Definitions

Harm
Poor patient outcome resulting from medical care rather than the natural history of the disease, whether or not it was preventable. This term includes adverse medical events (ie, falls, nosocomial infections), adverse drug events and adverse surgical events (ie, postoperative infections, surgical complications). It excludes medical errors that did not result in injury to patients.

Automated harm-detection method
A method of rapidly searching a large number of patient medical records with a computerised tool to identify actual harm, or indicators (associations) of harm. Records and events identified through computerised screening may then be subjected to further scrutiny by electronic or manual means to verify harm. We defined two degrees of automation: (1) fully automated methods, in which identification of harm was not followed by further chart review, and (2) partially automated methods, in which identified patient records were manually reviewed to verify harm.

Gold standard chart review
Manual review of the medical record initially by trained personnel, with subsequent review by either a physician or clinical pharmacist to confirm the presence or absence of harm and characteristics of such harm.

Data extraction and analysis

We developed and tested a standardised data form and extracted the following variables from included articles: details of patient sample, methodology used for automated harm detection, nature of events identified, description of alternative method of harm detection and comparisons of events detected by automated and alternative methods. Data were extracted by MG and AVC, with uncertainties resolved by discussion and consensus.

We critically appraised each study that compared the automated method of harm detection to a gold standard chart review using published criteria for validity of diagnostic test studies.[7] We assessed each study for: (a) independent, blind comparison of the automated method with a gold standard method, (b) performance of the gold standard assessment regardless of the automated method's results and (c) validation of the assessment in a second, independent set of patients.

If studies provided adequate data, we independently calculated the sensitivity, specificity and positive and negative predictive values of the automated harm-detection method.

RESULTS

Selection of articles

One hundred and seventy-six articles were reviewed for potential inclusion, of which 43 provided information on validity of automated methods of harm detection.[8–50] The remaining articles were excluded because they: were review articles on harm-detection methodologies (n=9)[51–59]; did not focus on detection of harm (n=26) or automated methods (n=22); did not include a comparison group (n=17); were not limited to inpatients (n=13); were descriptive papers of a program, incident reporting system, algorithm or computer simulation (n=33); were commentaries or editorials (n=11); or were repeat publications (n=2).

The methodologies and results from the 43 included studies are described in online appendix 1. Of these, 14 studies compared the automated harm-detection methodology to a gold standard chart review, and their methods and results are summarised in tables 1, 2.

As shown in online appendix 1, 20 studies were conducted among adult populations, three in paediatric patients, two among all age groups, one in geriatric patients, one among Medicare beneficiaries and one among patients 14 years and older. The most common hospital settings were general medical units (n=14), followed by general surgical units (n=8), medical, surgical or general intensive care units (n=8), medical subspecialties (n=3), neonatal and paediatric intensive care units (n=3) and obstetric units (n=2). The target population and setting were unstated in 15 studies.

Data sources for automated harm-detection methods

Automated harm-detection methods were classified into field-defined and natural language-processing systems. Field-defined systems relied on computerised detection using pre-existing numeric or coded data stored in medical records. Natural language processing relied on computerised analysis of free text within a medical record to detect language indicative of harm. Field-defined and natural language-processing systems are described in table 3.

Forty-one of 43 studies used field-defined systems for automated harm detection. The nature of the programs, databases used, data fields used and types of harm detected within this category were source-specific. Typical sources of data for field-defined programs included laboratory, radiology, microbiology, pharmacy, and administrative and billing databases. Five of 43 studies used natural language-processing systems. The most common source of data was discharge summaries. Radiology reports, chart text, daily progress notes, consultation notes, nursing records, and procedure or operative reports also were used.

Degree of automation

Twenty-five studies (58%) reported on detection tools that were partially automated,[8–14 21–25 31 32 34–38 40 45–48 50] 14 studies (33%) described fully automated tools,[15–17 19 26–30 33 41 42 44 49] and one study (2%) reported both fully and partially automated systems.[20] The degree of automation was unclear in three reports (7%).[18 39 43]

Types of events identified

Automated methods for detecting harm predominantly focused on identification of adverse drug events (ADEs) (n=21, 49%).[11 12 18 21–26 29–32 35–38 43 45 50] Ten automated methods (23%) focused on general adverse events,[8–10 19 33 34 40 46–48] eight (19%) focused on nosocomial infection,[14 20 28 39 41 42 44 49] and four (9%) focused on other specific adverse events (eg, decubitus ulcers, surgical complications).[13 15 17 27]

Accuracy of automated harm-detection methods

Only 14 studies[15 17 18 20 22 23 26 30 32–34 44 47 48] compared an automated harm-detection method with 'gold-standard' adverse event detection and were eligible for critical appraisal of validity (table 2). Methodologies used to evaluate these automated systems were heterogeneous. Seven studies (50%) applied the gold standard using independent, blind evaluators. Eight studies (57%) applied the gold standard independently of the outcome from the automated method. One study (7%) validated the results of the automated method in an independent, second set of patients.

Table 4 shows the sensitivity, specificity, and positive and negative predictive values of the automated methods that were

Qual Saf Health Care 2010;**19**:e11. doi:10.1136/qshc.2009.033027

Table 1 Summary of studies comparing automated harm-detection methods with gold standard chart review

Reference	Patient sample and time frame	Sampling strategy*	Specialty	Events identified	Automated event dataset sample size	Comparison event dataset sample size
Field-defined						
Nebeker *et al*[18]	Adults 2001 and 2003	Random	Unknown	Adverse drug events	3987 admissions	3987 admissions
Zhan *et al*[17]	Medicare benefic. 2002 to 2004	Random	General Surgery	Adverse event: specifically postoperative deep venous thrombosis and/or pulmonary embolism	20 868 hospital discharges identified as surgical patients	20 868 hospital discharges identified as surgical patients
Brossette *et al*[44]	Unknown 1–3 Dec 2003 and 26–29 Apr 2004	Sequential	Unknown	Infection	907 admissions	907 admissions
Hougland *et al*[30]	Adults 1 Jan 2001 to 31 Dec 2001	Random, Flagged sample (from records with at least one flagged adverse drug event code)	Unknown	Adverse drug events	3103 inpatients: 1961 random, 1142 flagged	Unknown
Polancich *et al*[15]	Unknown	Unknown	Unknown	Hospital acquired decubitus ulcers	Unknown	123 charts from patients with PSI-identified decubitus ulcers
Dormann *et al*[26]	Adults 1 Sept 2000 to 28 Feb 2001	Sequential	Gastroenterology	Adverse drug events	474 admissions of 377 patients; 109 ADEs	474 admissions of 377 patients; 109 adverse drug events
Trick *et al*[20]	Adults 1 Sept 2001 to 28 Feb 2002	Sequential	Unknown	Infection	135 positive blood cultures	144 positive blood cultures
Levy *et al*[23]	All age groups 1 Apr 1997 to 31 May 1997	Sequential	General Medical	Adverse drug events	199 admissions (192 patients)	199 admissions
Azaz-Livshits *et al*[22]	All age groups 1 Apr 1995 to 31 May 1995	Sequential	General Medical	Adverse drug events	153 admissions	153 admissions
Jha *et al*[32]	Adults 1 Oct 1994 to 31 May 1995	Sequential	MICU, SICU, General Medical, General Surgical	Adverse drug events	21 964 patient-days	21 964 patient-days
NLP						
Penz *et al*[47]	Adults 1 Jun 1999 to 31 Dec 2004	Sequential	MICU, SICU and other (placement of CVC)	Adverse events related to central venous catheter placement	316 patient records	40 patients records (10 very low probability† records, 30 high probability)
Forster *et al*[34]	Adults FY 2002	Random	General Medical, General Surgical	Adverse event	245 patients	245 patients
Melton and Hripcsak[48]	Unknown 1996–2000	Random (charts), Sequential (electronic discharge summaries)	Unknown	Adverse events: specifically 45 NYPORTS event types	1000 charts, 57 422 electronic discharge summaries	1000 charts
Murff *et al*[33]	Adults 1 Jan 2000 to 30 Jun 2000	Random (Cohort 1), Sequential (Cohort 2)	General Medical, Medicine subspecialties	Adverse drug events, adverse events, diagnostic errors, operative complications, falls	Cohort 1: 424 admissions. Cohort 2: 2826 admissions	Cohort 1: 295 Cohort 2: 145 Complex sampling/ subsampling and manual review process

*Sampling strategy refers to the method by which charts were chosen to be screened by the automated tool. Unless specifically noted, the same sampling strategy also applies to the gold standard method.
†A scoring system was developed by the authors to reflect the probability of the adverse event in question relating to the central venous catheter placement. This system is described in the text and in table 1 of the paper.
CVC, central venous catheter; MICU, Medical Intensive Care Unit; NYPORTS, New York Patient Occurrence Reporting and Tracking System; PSI, patient safety indicators; SICU, Surgical Intensive Care Unit.

Table 2 Evaluation of validity of studies comparing automated method to gold standard chart review

Reference	Strategy of event identification	Degree of automation*	Automated method source of data	Comparison method source of data	Gold standard applied by independent, blind reviewer?	Gold standard applied regardless of automated outcome?	Study method applied to independent patient set?	Comments
Field-defined								
Nebeker *et al*[18]	Computer algorithms	Chart review for study, unclear if strategy aims to be Full or Partial	ICD-9 CM codes	Medical record	Yes	Yes	No	Study used Hougland *et al*[30] methodology to specifically apply HOCTA (hierarchically optimal classification tree analysis) to administrative data to develop surveillance rules for the identification of ADEs manifesting as either bleeding or delirium. Requires expert computer programming.
Zhan *et al*[17]	Patient Safety Indicators	Full	ICD-9 CM codes	Medical record	Unknown	Unknown	No	DVT/PE events flagged by ICD-9 CM codes were compared with those discovered by gold standard chart review. The sample studied was a random sample abstracted by the Medicare Patient Safety Monitoring System.
Brossette *et al*[44]	Nosocomial Infection Marker	Full	Medical record and Lab database	Medical record	Yes	Yes	No	Nosocomial Infection Marker (NIM) program by Med Mined, Birmingham, Alabama. Took about 10 min/week to maintain. Total time for NIM: 2 h/10 000 admissions, compared with medical record review at 1.5 full time employees per 10 000 admissions.
Hougland *et al*[30]	Automated ICD-9 code strategy	Full: Review of flagged charts here for study purposes	ICD-9 CM codes	Medical record	Yes	Yes	No	Expert panel identified 416 ICD-9 CM codes to represent ADEs (flagged ADEs). Then chart review performed to ascertain codes' ability to detect/identify ADE.
Polancich *et al*[15]	Patient Safety Indicators	Full	Administrative data, Billing data, ICD-9 CM diagnosis and procedure codes	Medical Record	No	No	No	Designed to test validity of Agency for Healthcare Research and Quality (AHRQ) PSIs for detecting hospital acquired decubitus ulcers. Only a sample of cases was manually reviewed.
Dormann *et al*[26]	Automated laboratory signal detection	Full	Demographics, History, Lab findings, Drugs, & Diagnosis	Medical record	Unknown	Unknown	No	Used automated lab signals (ALS) and changes in ALS to identify ADEs. Automated system used to flag potential ADEs, which were then sent as an alert to physicians. Use of delta ALS (change) resulted in improvement over Dormann *et al*'s[25] methodology.
Trick *et al*[20]	Computer algorithm	Full and Partial	Medical record; Lab, pharmacy, & radiology database; Microbiology	Medical record; Lab, pharmacy, & radiology database; Microbiology	Yes	Yes	No	Comparison of manual and computer assisted bloodstream central venous catheter infection surveillance using data from two hospitals. Different computer algorithms developed for full or partial automation were tested.
Levy *et al*[23]	Automated laboratory signal detection	Partial	Lab database	Lab database and clinical data	Unknown	Yes	No	Implementation of the pilot program described in Azaz-Livshits *et al*.[22] Computerised lab data monitored to detect ADEs using the same signals as the pilot study.
Azaz-Livshits *et al*[22]	Automated laboratory signal detection	Partial	Lab database	Lab database and clinical data	Unknown	Yes	No	Pilot program to develop and assess computerised laboratory data as a detection tool for ADE in 34-bed medical ward in Jerusalem, Israel. Lab signals generated by computer, then verified by team. Limited computerised patient data at this hospital; however lab data were fully electronic. Cost of this system reasonable compared with costs of ADEs.

Continued

Table 2 Continued

Reference	Strategy of event identification	Degree of automation*	Automated method source of data	Comparison method source of data	Gold standard applied by independent, blind reviewer?	Gold standard applied regardless of automated outcome?	Study method applied to independent patient set?	Comments
Jha *et al*[32]	Automated triggers	Partial	Medical record	Medical record	Yes	Unknown	No	Study of computer-based ADE identification using modified Classen 1991[8] rules to create automated triggers with which the electronic record was screened. Rules modified during the study to increase PPV, and new rules created. Trained reviewer and physician were blinded to detection method. 11 person-hours per week for automated method versus 55 for chart review and 5 for voluntary reporting.
NLP								
Penz *et al*[47]	Computer algorithms & Natural Language Processing	Partial	Text records: Daily progress notes; Consultation, Nursing, and Procedure notes; Operative reports; Discharge summaries	Text records: Daily progress notes; Consultation, Nursing, and Procedure notes; Operative reports; Discharge summaries	No	No	No	Compared two methods for semiautomated review of text records within the VA database using NLP (MedLEE) and a phrase matching algorithm (PMA). Limited by incomplete or inaccurate documentation, incomplete coding, spelling errors, and sentence structure abbreviations. Time/technology intensive.
Forster *et al*[34]	Computerised screen for trigger words in free text	Partial	Discharge summaries	Discharge summaries	Yes	Yes	No	Automated adverse event lexicon made up of 104 terms used by Murff *et al*.[33] Computerised search engine scanned discharge summaries (dtsearch desktop) to detect potential harm. Specificity higher for non-elective admissions and discharge summaries dictated by residents/staff versus medical students. Automated detection reduced physician time by one-fifth.
Melton and Hripcsak[48]	Natural Language Processing	Partial	Discharge summaries	Full electronic chart; combined electronic chart and paper chart for a subset of 100 patients	No	No	No	Natural Language Processing system (MedLEE) to identify 45 NY Patient Occurrence Reporting and Tracking System event types. Chart review by physician and independent informatician of random sample of 1000 charts to assess performance of NLP program. Results biased towards patients with electronic discharge summaries. This method is technologically intensive.
Murff *et al*[33]	Computerised screen for trigger words in free text	Full (goal is a fully automated system, manual review of subsamples performed for study)	Discharge summaries	Medical record (not otherwise specified)	Yes	Yes†	Yes	Brigham and Women's Hospital, using Brigham Integrated Computer system. Computerised screening tool searched free text discharge summaries for trigger words indicating possible adverse events. List of automated trigger words compiled using Harvard Medical Practice Study definitions as base. Electronic method alone versus electronic plus manual review compared for two cohorts. Reviewers blinded to whether screening tool had identified the admission.

ADE, adverse drug event; AE, adverse event; DVT, deep venous thrombosis; NLP, natural language processing; PE, pulmonary embolism; PPV, positive predictive value; PSI, patient safety indicators; VA, Veterans Administration.
*We define fully automated methods as those where the identification of harm was not followed by further chart review, and partially automated methods where patient records flagged by the automated detection of potential harm (eg, 'trigger') were manually reviewed to verify harm.
†Authors manually reviewed a random 25% sample of screened-negative charts, then used this random sample to estimate the number of adverse events occurring in entire set of screened-negative charts.

Table 3 Description and classification of field-defined and natural language processing systems for automated detection of harm*

Automated method	Data source used	Events identified	Comments
Complications Screening Program (CSP)[8–10 46]	ICD-9 CM codes	Adverse drug events, adverse surgical outcomes, infections, and miscellaneous complications such as falls	A computerised method for identifying potentially preventable complications of hospital care.
Health Evaluation through Logical Processing (HELP)[11–14]	Electronic Medical Record: specifically including pharmacy, laboratory, radiology and surgery records	Adverse drug events, adverse medical device events, infection	Integrated electronic medical record of the LDS Hospital in Salt Lake City, Utah, which contains an interactive modular knowledge base that continually analyses information
Patient Safety Indicators (PSI)[15–17 46]	Administrative data: billing information, ICD-9 CM diagnosis codes and procedure codes	Adverse events	A fully automated method developed by the Agency for Healthcare Research and Quality
Computer algorithms[18–21]	Electronic Medical Record: components specific to the particular program: see online appendix 1	Adverse events, adverse drug events, infection	Specific, named computer programs
Lab signal detection tools[22–26]	Laboratory Database	Adverse drug events	Automated tools search for key words or word combinations that signal potential or actual harm—for example, detection of elevated potassium levels
ICD-9 CM or billing code detection tools[27–30]	Administrative data: ICD-9 CM or billing codes	Adverse drug events, infections, surgical complications	Automated tools scan for diagnosis, discharge, or billing codes that signal potential or actual harm—for example, evidence of antibiotic exposure following a postoperative infection
Tools using computerised triggers[31–45 50]	Electronic Medical Record: multiple sources such as pharmacy, laboratory, and microbiology databases	Adverse events, adverse drug events infection	Automated tools using multiple triggers to signal actual or potential harm—for example, detection of elevated potassium levels (laboratory database) combined with certain medication administration (pharmacy database). Among the various tools included in this category, there are four named systems: Dynamic Pharmaco-Monitoring System, Nosocomial Infection Marker, Event Detector, New York Antimicrobial Resistance Project.
Natural language processing systems[33 34 47–49]	Free text in the Electronic Medical Record: discharge summaries, radiology reports, chart notes	Adverse events, infection	Sophisticated programs that 'read' free text via the application of computer logic

*Multiple detection strategies were used in several studies, including those that combined two or more field-defined systems,[46] two natural language-processing systems,[47] and both a field-defined and natural language-processing system.[33 34]

Table 4 Accuracy of automated methods for event identification*

Reference	Events identified by automated harm-detection method	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)
Field-defined					
Nebeker *et al*[18]	Calculated separately for bleeding/ anticoagulation ADEs and delirium ADEs	Bleeding: 0.86† Delirium: 0.94†	Bleeding: 0.89† Delirium: 0.71†	Bleeding: 0.12† Delirium: 0.03†	NA
Zhan *et al*[17]	DVT Cases	0.67 (0.58 to 0.76)	NA†	0.31 (0.25 to 0.37)	NA†
	PE cases	0.74 (0.59 to 0.90)	NA†	0.24 (0.16 to 0.33)	NA†
	DVT/PE Cases	0.68 (0.60 to 0.76)	NA†	0.29 (0.24 to 0.34)	NA†
Brossette *et al*[44]	Hospital-wide nosocomial infection	0.88†¶	NA†¶	0.78	NA
Hougland *et al*[30]	Codes for inpatient ADE	0.10 (0.63 to 0.14)	0.97 (0.96 to 0.98)	0.32 (0.22 to 0.43)	0.89 (0.88 to 0.91)
Polancich *et al*[15]	Patients with decubitus ulcers	NA†	NA†	0.50 (0.42 to 0.59)	NA†
Dormann *et al*[26]	ADR positive admissions using NEW ALS	0.91	0.23	0.18‡	0.93
	ADR positive admissions using DELTA ALS	0.41	0.76	0.25‡	0.87
Trick *et al*[20]	Hospital-acquired episodes of primary CVC associated bloodstream infections	0.81†	0.72†	0.62†	0.87†
Levy *et al*[23]	Admissions	0.63 (0.51 to 0.74)	0.42 (0.34 to 0.51)	0.34 (0.25 to 0.42)	0.70 (0.60 to 0.80)
Azaz-Livshits *et al*[22]	Admissions	0.66 (0.51 to 0.81)	0.51 (0.42 to 0.60)	0.31 (0.21 to 0.41)	0.82 (0.73 to 0.91)
Jha *et al*[32]	ADE	NA†	NA†	0.16§ (0.16 to 0.19)	NA†
NLP					
Penz *et al*[47]	Cases	PMA: 0.70 † NLP: 0.50 † Combination: 0.72†	PMA: 0.55 † NLP: 0.91 † Combination: 0.80†	PMA: 0.41† NLP: 0.71† Combination: 0.64†	PMA: 0.8† NLP: 0.8† Combination: 0.85†
Forster *et al*[34]	Patients	0.23 (0.11 to 0.35)	0.92 (0.88 to 0.96)	0.41 (0.22 to 0.59)	0.83 (0.78 to 0.88)
Melton *et al*[48]	Cases	0.28 (0.16 to 0.40)	0.98 (0.97 to 0.99)	0.47 (0.30 to 0.64)	0.96 (0.95 to 0.97)
Murff *et al*[33]	AE	Fully automated: 0.69 (0.62 to 0.75) Partially automated: 0.64 (0.56 to 0.70)	Fully automated 0.48 (0.42 to 0.55) Partially automated: 0.85 (0.80 to 0.90)	Fully automated: 0.52 (0.46—0.58) Partially automated: 0.78 (0.72—0.85) (cohort 1), 0.84 † (cohort 2)	Fully automated: 0.65 (0.58—0.72) Partially automated: 0.74 (0.69—0.79)

*95% CIs for independently verified values reported in parentheses.
†Denotes figures that we could not independently verify.
‡Dormann *et al*[26] defined the positive predictive value (PPV) as the number of alerts associated with adverse drug reactions (ADRs) out of the total number of alerts. Using this criteria, they found the following PPVs: New automatic laboratory signals (ALS) (574/2328) 25%; Delta ALS (189/580) 32%.
§Jha *et al*[32] report a range of PPVs based on the first and final 8 weeks of data collection (0.16 and 0.23, respectively). We were able to independently verify the PPV for the first 8 weeks of the study only.
¶Brossette *et al*[44] reported a sensitivity of 0.86 and a specificity of 0.98. It is unclear how they identified true negative screens.
ADE, adverse drug event; AE, adverse event; CVC, central venous catheter; DVT, deep venous thrombosis; NLP, natural language processing; PE, pulmonary embolism; PMA, phrase matching algorithm.

compared against a gold standard chart review. Sensitivities of different methods ranged from 0.10 to 0.94, and specificities ranged from 0.23 to 0.98. Positive predictive values ranged from 0.03 to 0.84, and negative predictive values ranged from 0.70 to 0.96. Our independent assessment of validity allowed us to verify all published values for nine of the 14 studies that reported validity data.[15 17 22 23 30 33 34 48] Figure 1 displays the sensitivity and 1-specificity intersection points of methods used in these studies in a format similar to that of a receiver-operating characteristic curve.

DISCUSSION

Strategies to improve patient safety require efficient and accurate detection of patient harm. Automated methods of harm detection have been used for this purpose because they offer the potential to rapidly scan patient records with minimal human effort. This systematic review describes types of automated methods of harm detection used in inpatient settings, events identified by these methods and their accuracy.

We found two categories of automated harm detection described in the literature: field-defined systems (used in most studies) and natural language-processing systems. Most frequently laboratory, pharmacy and administrative databases were used to identify adverse drug events, general adverse events and nosocomial infections.

We found that the validity of studies describing automated harm-detection methods was variable. Of these studies, those attempting to identify ADEs [18 30] and nosocomial infections[20 44] using field-defined methods, and one attempting to identify multiple types of adverse events[33] using natural language processing satisfied more validity criteria than others. We believe that automated harm-detection methods will have more validity if they attempt to identify events that are discrete, easily and reliably detected, and consistently documented in the chart, such as adverse drug events, nosocomial infections, pressure ulcers and postoperative complications.

Automated harm detection has the potential to positively impact clinical practice. While most automated methods retrospectively identified harm, eight were paired with real-time surveillance alerts that informed physicians or pharmacists of an adverse event. Such prospective surveillance systems can alert the clinical team of impending or ongoing harm, thus allowing early intervention to limit harm. Real-time alerts were present within methods for detecting adverse drug events,[11 21 23 26 35 45] general adverse events[40] and nosocomial infection.[14] Automated alerts were a component of the Health Evaluation through Logical Processing system[11 14] and were incorporated within methods using automated lab signal detection,[23 26 45] computer algorithms[21] and other automated triggers.[35 40]

Another potential benefit of automated detection is the reduction of person-hours required for harm surveillance. Few studies[14 21 22 32 34 38 40 44] provided information on financial or human resource requirements for implementing and maintaining automated detection tools. In general, the automated methods reviewed here require fewer person-hours than manual

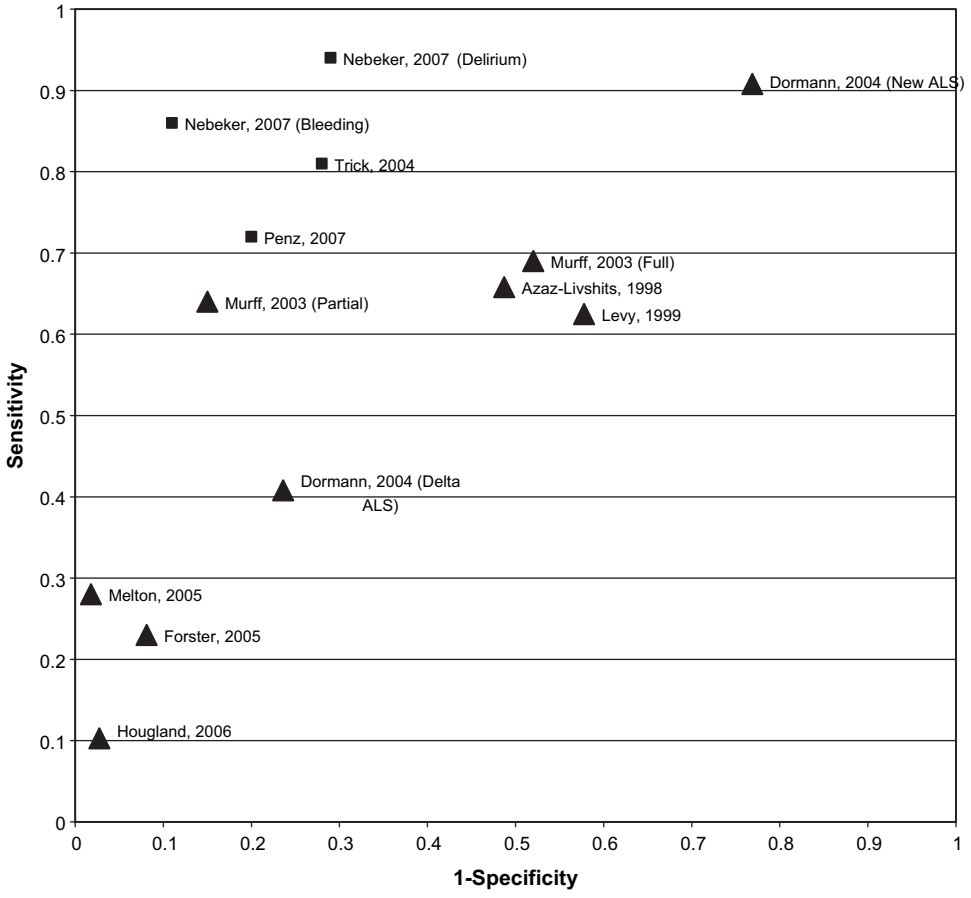

Figure 1 Sensitivity by 1-specificity for automated methods compared with gold standard methods of harm detection.

[a] Only the aggregated values for the harm detection method from each paper are shown. Individual components of an automated method are not shown.

[b] Triangles represent sensitivity and specificity values that could be independently verified.

chart review. Field-defined strategies appear to be less technologically demanding than natural language-processing strategies. Sophisticated computer algorithms and natural language-processing programs require specialised subject knowledge, skill and time to develop, and require installation and instruction by experts.[18 48] Whether costs to implement such programs are offset by savings from eliminating manual chart review and decreased patient harm is unknown and should be studied. Future studies also should quantify differences in time and personnel resources needed for the automated detection method, relative to other detection strategies.

To our knowledge, four of the 43 unique articles report on commercially available automated harm-detection systems (MedLEE,[48] dtsearch desktop,[34] Nosocomial Infection Marker (NIM)[44] and Dynamic Pharmaco-Monitoring System[45]). Other articles report on systems that employ data elements common across medical institutions (ie, ICD-9 codes used in the Complications Screening Program[8–10]) use software available to the VA or specific states (ie, RADARx, NY Antimicrobial Resistance Project[21 42]) or are available through the Agency for Healthcare Research and Quality (ie, Patient Safety Indicators[15–17]). The availability of the remaining detection systems is either institution-specific or not made clear by their developers.

While automated tools offer promise for efficient and accurate harm detection, there are important limitations that currently make them unsuitable for widespread application, particularly for interhospital comparisons. The reported sensitivity and specificity are variable and often low, suggesting that many episodes of harm may go undetected, and that many events identified will be false positives. Low accuracy may result from limited capability of the tool to detect events, or from flawed sources of data used for automated harm detection. For example, the reliability of field-defined systems can be affected by data entry errors or limited availability and accuracy of administrative codes, while natural language processing is sensitive to spelling and grammatical errors in free text. Both systems may include irrelevant or erroneous information, or exclude necessary information. For example, perhaps driven by medical-legal concerns, health professionals often do not include information about medical errors and resulting adverse events in their progress notes, problem lists and discharge summaries. Thus, an electronic medical record containing accurate, complete and easily accessible information can enhance the performance of an automated detection tool. Understanding these factors is important when evaluating the technological requirements, feasibility and inherent limitations of automated detection methods.

The variety of distinct automated methodologies makes comparisons between studies and between automated tools difficult and unreliable. Differences in the quality and content of data sources, as well as other unknowns such as accuracy of hospital documentation and coding practices, also complicate comparisons. The performance and methods of automated tools also may be institution-specific, making it difficult to generalise to other organisations or patient populations. For example, the Health Evaluation through Logical Processing system used by LDS Hospital in Salt Lake City, Utah relies on an advanced, highly integrated and dynamic information system that is not widely available.[11–14]

We speculate that field-defined methods of automated harm detection will prove superior to natural language-processing methods, particularly if information about harm is accurately documented in electronic medical record systems in prespecified fields, thus allowing rapid and reliable detection of harm events.

The methodological rigour of studies was variable. Only two-thirds of the 14 studies that compared an automated method with a gold standard chart review had verifiable validity results. Moreover, most studies compared automated harm-detection methods with other sources of data on patient harm (eg, voluntary reporting,[11–13 24 25 29 31 37 38 50] unstandardised chart reviews,[8 10 14 28 36 41 43 45] and prospective surveillance records[42 49]). The validity of data from studies without chart review comparison is questionable given the absence of a defined denominator of events against which to measure the performance of the automated tool. The use of different methods, statistical analyses, denominator values and outcomes precludes a comparison of one automated method with another, as well as any attempt to statistically pool their results in a meta-analysis.

Other authors have summarised the literature on automated harm-detection methods, but most have focused on automated methods specific to a type of harm (ie, adverse drug events[51 54] or nosocomial infections),[59] patient population (ie, paediatrics),[52] source of data (ie, administrative data)[57] or automated technology (ie, natural language processing).[58] Our systematic review included all types of automated methods, harm events and sources of data evaluated in an inpatient setting. Furthermore, we provide an additional level of critical appraisal compared with other systematic reviews.[55 56] For example, while Bates *et al*[55] address differences between study methodologies by noting the presence or absence of gold standard comparison, they do not assess validity of studies or independently verify reported data. To our knowledge, this is the first systematic review to critically assess methodological rigour and study validity.

While our review has several strengths, it also has limitations. First, the search strategy was limited to published English language articles. Second, we did not evaluate scientific meeting abstracts, nor did we contact investigators to identify unpublished studies. Third, publication bias must be considered in which studies with negative findings may not have reached dissemination venues. Fourth, most of the articles evaluated automated methods of harm detection among adults in general medical or surgical units, which may limit application to other populations and settings. Finally, our independent appraisal of the methodology and validity of key studies relied on information available within published articles. Our inability to verify the rigour and validity of all studies highlights the variation among even the most rigorous evaluations.

In conclusion, our review identified numerous automated methods of harm detection in two broad categories—field-defined methods and natural language processing—that identified a broad range of harm events, but particularly adverse drug events and nosocomial infections. Although many of these studies described the accuracy (sensitivity and specificity) of automated harm detection when compared with chart review, these results may not be valid due to methodological flaws in the conduct of many of these studies. Future studies assessing the performance of automated harm-detection methods should ensure that the gold-standard assessment (usually chart review) is performed by a blinded assessor, the gold-standard is applied independently of the results of the automated method (ie, charts not flagged by the automated method are reviewed for false negatives), and the automated method is tested in a set of patients that is independent of the set used to develop the automated method. Finally, efforts should be made to improve documentation of harm episodes in the patient record, in problem lists and when generating diagnosis codes, in order to

improve automated harm detection. Future research should also focus on developing methods for real-time harm detection. In this way, automated harm-detection tools will realise their potential to describe accurately the incidence of harm in hospitalised patients, monitor changes from preventive interventions, and compare institutions and individual health professionals. Establishing universal standards and guidelines for the development, testing and utilisation of automated harm-detection methods, perhaps through a centralised agency, would allow data to be collected and compared in a rigorous, systematic fashion.

Summary

Automated methods of harm detection are feasible, allow rapid scanning of a large number of patient records with minimal effort and have the potential to identify events as they occur or soon thereafter. However, the heterogeneity of automated methodologies, the spectrum of study rigour and the widely varying accuracy data suggest that currently available automated methods poorly measure the true incidence of harm. These methods cannot replace chart review as the gold standard but can provide estimates of the frequency of harm that can allow hospitals to identify priorities for action, make decisions about safety interventions and potentially monitor change over time. As automated harm-detection tools and scientific methods to test them evolve, there exists a great potential to positively impact patient safety.

Acknowledgements We are grateful for the administrative support provided by the Institute for Healthcare Improvement.

Funding Funding for the literature review was provided by the Institute for Healthcare Improvement (IHI) to MG and ADVC. Subsequent data analysis and interpretation, as well as conceptualisation, preparation, and review of the manuscript were not financially supported.

Competing interests JK-C was employed by Premier Inc. from 31 March 2007 to 2 July 2008. Premier has developed an automated event detection product, SafetySurveillor. This study does not reference or endorse this product. No other authors disclosed any potential conflicts of interest.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Institute of Medicine.** *To err is human*. Washington, DC: National Academy Press, 1999.
2. **Institute for Healthcare Improvement.** Available at: http://www.IHI.org (accessed 14 Jan 2008).
3. **Thomas EJ,** Studdert DM, Burstin HR, *et al*. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;**38**:261—71.
4. **Brennan TA,** Leape LL, Laird NM, *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.
5. **Leape LL,** Brennan TA, Laird N, *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.
6. **Wilchesky M,** Tamblyn R, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;**57**:131—41.
7. **Straus SE,** Richardson WS, Glasziou P, *et al*. *Evidence-based medicine: how to practice and teach EBM*. 3rd edn. Edinburgh: Churchill Livingstone, 2005.
8. **Iezzoni LI,** Foley SM, Heeren T, *et al*. A method for screening the quality of hospital care using administrative data: preliminary validation results. *QRB Qual Rev Bull* 1992;**18**:361—71.
9. **Weingart SN,** Iezzoni LI, Davis RB, *et al*. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care* 2000;**38**:796—806.
10. **Lawthers AG,** McCarthy EP, Davis RB, *et al*. Identification of in-hospital complications from claims data. Is it valid? *Med Care*2000;**38**:785—95.
11. **Classen DC,** Pestotnik SL, Evans RS, *et al*. Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991;**266**:2847—51.
12. **Evans RS,** Pestotnik SL, Classen DC, *et al*. Development of a computerized adverse drug event monitor. *Proc Annu Symp Comput Appl Med Care* 1991:23—7.
13. **Samore MH,** Evans RS, Lassen A, *et al*. Surveillance of medical device-related hazards and adverse events in hospitalized patients. *JAMA* 2004;**291**:325—34.
14. **Evans RS,** Larsen RA, Burke JP, *et al*. Computer surveillance of hospital-acquired infections and antibiotic use. *JAMA* 1986;**256**:1007—11.
15. **Polancich S,** Restrepo E, Prosser J. Cautious use of administrative data for decubitus ulcer outcome reporting. *Am J Med Qual* 2006;**21**:262—8.
16. **McDonald KM,** Romano PS, Geppert J, *et al*. *Measures of Patient Safety Based on Hospital Administrative Data—The Patient Safety Indicators. Technical Review Number 5. (Prepared by the University of California San Francisco-Stanford Evidence-based Practice Center under Contract No. 290-97-0013). AHRQ Publication No. 02-0038.* Rockville, MD: Agency for Healthcare Research and Quality; August 25 2002.
17. **Zhan C,** Battles J, Chiang YP, *et al*. The validity of ICD-9-CM codes in identifying postoperative deep vein thrombosis and pulmonary embolism. *Jt Comm J Qual Patient Saf* 2007;**33**:326—31.
18. **Nebeker JR,** Yarnold PR, Soltysik RC, *et al*. Developing indicators of inpatient adverse drug events through nonlinear analysis using administrative data. *Med Care* 2007;**45**(10 Suppl 2):S81—8.
19. **Benson M,** Junger A, Fuchs C, *et al*. Using an anesthesia information management system to prove a deficit in voluntary reporting of adverse events in a quality assurance program. *J Clin Monit Comput* 2000;**16**:211—17.
20. **Trick WE,** Zagorski BM, Tokars JI, *et al*. Computer algorithms to detect bloodstream infections. *Emerg Infect Dis* 2004;**10**:1612—20.
21. **Brown S,** Black K, Mrochek S, *et al*. RADARx: recognizing, assessing, and documenting adverse Rx events. *Proc AMIA Symp* 2000:101—5.
22. **Azaz-Livshits T,** Levy M, Sadan B, *et al*. Computerized surveillance of adverse drug reactions in hospital: pilot study. *Br J Clin Pharmacol*1998;**45**:309—14.
23. **Levy M,** Azaz-Livshits T, Sadan B, *et al*. Computerized surveillance of adverse drug reactions in hospital: implementation. *Eur J Clin Pharmacol* 1999;**54**:887—92.
24. **Bagheri H,** Michel F, Lapeyre-Mestre M, *et al*. Detection and incidence of drug-induced liver injuries in hospital: a prospective analysis from laboratory signals. *Br J Clin Pharmacol* 2000;**50**:479—84.
25. **Dormann H,** Muth-Selbach U, Krebs S, *et al*. Incidence and costs of adverse drug reactions during hospitalisation: computerised monitoring versus stimulated spontaneous reporting. *Drug Saf* 2000;**22**:161—8.
26. **Dormann H,** Criegee-Rieck M, Neubert A, *et al*. Implementation of a computer-assisted monitoring system for the detection of adverse drug reactions in gastroenterology. *Aliment Pharmacol Ther* 2004;**19**:303—9.
27. **Roos LL Jr,** Cageorge SM, Austen E, *et al*. Using computers to identify complications after surgery. *Am J Public Health* 1985;**75**:1288—95.
28. **Hirschhorn LR,** Currier JS, Platt R. Electronic surveillance of antibiotic exposure and coded discharge diagnoses as indicators of postoperative infection and other quality assurance measures. *Infect Control Hosp Epidemiol* 1993;**14**:21—8.
29. **Seeger JD,** Schumock GT, Kong SX. Estimating the rate of adverse drug reactions with capture—recapture analysis. *Am J Health Syst Pharm* 1996;**53**:178—81.
30. **Hougland P,** Xu W, Pickard S, *et al*. Performance of international classification of diseases, 9th revision, clinical modification codes as an adverse drug event surveillance system. *Med Care* 2006;**44**:629—36.
31. **Whipple JK,** Quebbeman EJ, Lewis KS, *et al*. Identification of patient-controlled analgesia overdoses in hospitalized patients: a computerized method of monitoring adverse events. *Ann Pharmacother* 1994;**28**:655—8.
32. **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;**5**:305—14.
33. **Murff HJ,** Forster AJ, Peterson JF, *et al*. Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc* 2003; **10**:339—50.
34. **Forster AJ,** Andrade J, van Walraven C. Validation of a discharge summary term search method to detect adverse events. *J Am Med Inform Assoc* 2005;**12**:200—6.
35. **Hartis CE,** Gum MO, Lederer JW Jr. Use of specific indicators to detect warfarin-related adverse events. *Am J Health Syst Pharm* 2005;**62**:1683—8.
36. **McIntosh ST,** Petropoulos JB. Using data from automated dispensing units to identify adverse drug reactions. *Am J Health Syst Pharm* 2005;**62**:2397—400.
37. **Kilbridge PM,** Campbell UC, Cozart HB, *et al*. Automated surveillance for adverse drug events at a community hospital and an academic medical center. *J Am Med Inform Assoc* 2006;**13**:372—7.
38. **Kilbridge PM,** Alexander L, Ahmad A. Implementation of a system for computerized adverse drug event surveillance and intervention at an academic medical center. *J Clin Outcomes Manage* 2006;**13**:94—100.
39. **Pokorny L,** Rovira A, Martin-Baranera M, *et al*. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol* 2006;**27**:500—3.
40. **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;**15**:184—90.
41. **Bellini C,** Petignat C, Francioli P, *et al*. Comparison of automated strategies for surveillance of nosocomial bacteremia. *Infect Control Hosp Epidemiol* 2007;**28**:1030—5.
42. **Graham PL 3rd,** San Gabriel P, Lutwick S, *et al*. Validation of a multicenter computer-based surveillance system for hospital-acquired bloodstream infections in neonatal intensive care departments. *Am J Infect Control* 2004;**32**:232—4.
43. **Huang C,** Noirot LA, Reichley RM, *et al*. Automatic detection of spironolactone—related adverse drug events. *AMIA Annu Symp Proc* 2005:989.

44. **Brossette SE,** Hacek DM, Gavin PJ, *et al*. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? *Am J Clin Pathol* 2006;**125**:34—9.
45. **Seger AC,** Jha AK, Bates DW. Adverse drug event detection in a community hospital utilising computerised medication and laboratory data. *Drug Saf* 2007;**30**:817—24.
46. **Weissman JS,** Rothschild JM, Bendavid E, *et al*. Hospital workload and adverse events. *Med Care* 2007;**45**:448—55.
47. **Penz JF,** Wilcox AB, Hurdle JF. Automated identification of adverse events related to central venous catheters. *J Biomed Inform* 2007;**40**:174—82.
48. **Melton GB,** Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;**12**:448—57.
49. **Haas JP,** Mendonca EA, Ross B, *et al*. Use of computerized surveillance to detect nosocomial pneumonia in neonatal intensive care unit patients. *Am J Infect Control* 2005;**33**:439—43.
50. **Ferranti J,** Horvath MM, Cozart H, *et al*. Reevaluating the safety profile of pediatrics: a comparison of computerized adverse drug event surveillance and voluntary reporting in the pediatric environment. *Pediatrics* 2008;**121**:e1201—7.
51. **Handler SM,** Altman RL, Perera S, *et al*. A systematic review of the performance characteristics of clinical event monitor signals used to detect adverse drug events in the hospital setting. *J Am Med Inform Assoc* 2007;**14**:451—8.
52. **Jacobs B.** Electronic medical record, error detection, and error reduction: a pediatric critical care perspective. *Pediatr Crit Care Med* 2007;**8**(2 Suppl):S17—S20.
53. **Chaudhry B,** Wang J, Wu S, *et al*. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;**144**:742—52.
54. **Anderson JG.** Information technology for detecting medication errors and adverse drug events. *Expert Opin Drug Saf* 2004;**3**:449—55.
55. **Bates DW,** Evans RS, Murff H, *et al*. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;**10**:115—28.
56. **Murff HJ,** Patel VL, Hripcsak G, *et al*. Detecting adverse events for patient safety research: a review of current methodologies. *J Biomed Inform* 2003;**36**:131—43.
57. **Zhan C,** Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;**12**(Suppl 2):ii58—63.
58. **Spyns P.** Natural language processing in medicine: an overview. *Methods Inf Med* 1996;**35**:285—301.
59. **Leal J,** Laupland KB. Validity of electronic surveillance systems: a systematic review. *J Hosp Infect* 2008;**69**:220—9.