International Journal for Quality in Health Care 2010; **Volume 22, Number 4:** pp. 266–274
Advance Access Publication: 9 June 2010
10.1093/intqhc/mzq026

Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool

JAMES M. NAESSENS[1], THOMAS J. O'BYRNE[1], MATTHEW G. JOHNSON[1], MONICA B. VANSUCH[1], COREY M. MCGLONE[1] AND JEANNE M. HUDDLESTON[2]

[1]Division of Health Care Policy & Research, and [2]Department of General Internal Medicine, Mayo Clinic, Rochester, Minnesota

Address reprint requests to: James M. Naessens, Mayo Clinic, 200 First Street Southwest, Rochester, MN 55905, USA. Tel: +1-507-284-5592; Fax: +1-507-284-1739; E-mail: naessens@mayo.edu

Accepted for publication 2 May 2010

Abstract

Objective. To determine the inter-rater reliability of the Institute for Healthcare Improvement's Global Trigger Tool (GTT) in a practice setting, and explore the value of individual triggers.

Design. Prospective assessment of application of the GTT to monthly random samples of hospitalized patients at four hospitals across three regions in the USA.

Setting. Mayo Clinic campuses are in Minnesota, Arizona and Florida.

Participants. A total of 1138 non-pediatric inpatients from all units across the hospital.

Intervention. GTT was applied to randomly selected medical records with independent assessments of two registered nurses with a physician review for confirmation.

Main Outcome Measure. The Cohen Kappa coefficient was used as a measure of inter-rater agreement. The positive predictive value was assessed for individual triggers.

Results. Good levels of reliability were obtained between independent nurse reviewers at the case-level for both the occurrence of any trigger and the identification of an adverse event. Nurse reviewer agreement for individual triggers was much more varied. Higher agreement appears to occur among triggers that are objective and consistently recorded in selected portions of the medical record. Individual triggers also varied on their yield to detect adverse events. Cases with adverse events had significantly more triggers identified (mean 4.7) than cases with no adverse events (mean 1.8).

Conclusions. The trigger methodology appears to be a promising approach to the measurement of patient safety. However, automated processes could make the process more efficient in identifying adverse events and has a greater potential of improving care delivery and patient '*outcomes*'.

Keywords: medical errors, adverse events, reliability, hospitals

Introduction

Attempts to address the growing costs of health care in the USA have brought increased attention to the issue of safety in medical care, with policies aimed at reducing hospital complications, readmissions and adverse events. Actions have been taken by the Center for Medicare and Medicaid Systems to eliminate payment for selected hospital acquired conditions [1], while some states require that providers publicly report the incidence of the National Quality Forum's list of 28 'never' events [2]. Patient safety advocates have stressed that identifying and reporting adverse events with harm rather than all errors in care would lead to safer environments for patients [3].

The Institute of Medicine defines an adverse event as always associated with 'unintended harm to the patient by an act of commission or omission rather than by the underlying disease or condition of the patient' [4]. Not all adverse events are preventable, nor are they always the result of medical errors. Medical errors are mistakes or failures in the process of care. While they have the potential of being harmful, often they are not linked to patient injury [3]. Moreover, when medical errors do lead to adverse events, many are minor in terms of patient harm.

In December 2006, the Institute for Healthcare Improvement (IHI) leadership announced the '5-million lives campaign' to foster prevention of adverse events. More than 70% of US hospitals committed to active participation, including regular measurement and transparent reporting of their institution's adverse event rates [5]. Voluntary reporting is known to undercount the number of hospital adverse events [6]. Therefore, IHI Global Trigger Tool (GTT) was developed to provide a measurement tool to determine the degree of campaign success. While many hospitals in the USA are employing the GTT, there is a dearth of data regarding actual utility of this tool. The GTT appears to identify more events than other methods, but requires substantial manual effort. Using a combination of Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSI) and provider-reported events, we reported that 4% of hospital discharges had adverse events. However, when using the GTT on a sample of discharges, over 27% had adverse events [7].

In this study, we set out to evaluate three aspects of utility for the GTT: (1) inter-rater reliability for identifying triggers and adverse events in large academic institutions in geographically distinct areas of the country using the published IHI GTT methodology [8, 9]; (2) frequency of trigger identifications and (3) the predictive ability of the individual triggers to identify adverse events. Academic hospitals of Mayo Clinic in Arizona, Florida and Minnesota were included in this assessment. The project primarily involved physician and nurse reviewer efforts with additional support from research and information technology.

Methods

Setting

Mayo Clinic, existing in three main practices: Rochester MN, Scottsdale AZ and Jacksonville FL, is a large multispecialty academic health care institution which provides primary care, specialty and subspecialty care, including hospital services, to large patient populations in three geographic regions as well as to nationally and internationally referred patients. This research was approved by the Institutional Review Board covering all three study sites.

Study population

Every 2 weeks at each of the three practices, 10 hospitalized patients were randomly selected utilizing electronic administrative databases. Only patients with complete charts (discharge summaries and coding completed) who had been discharged in the previous 14 days were eligible. Inpatient admissions were included from the psychiatric, physical rehabilitation, obstetric, general medical, specialty and surgical units. Pediatric patients were not included.

Study procedures

A brief review of the medical records of the sampled patients, lasting no more than 20 min, was conducted independently by each of two trained nurses available from a dedicated pool, for the presence of any of 55 'triggers' using the IHI GTT [9]. The GTT review form is attached as Appendix [10]. For the purposes of this study, 'triggers' were defined as sentinel conditions believed to be associated with the occurrence of an adverse event. The identification of a 'trigger' by each of the nurse reviewers led to a further, more exhaustive review of the care delivery near the time the 'trigger' occurred. The occurrence of an adverse event was determined and in cases where an adverse event was identified, the level of harm to the patient was assessed. A physician reconciled the independently recorded triggers and adverse events by each nurse, after assessing the presence of the identified triggers, occurrence of adverse events and the level of harm.

An example of the use of one such trigger is the presence in the patient's chart of an international normalized ratio (INR) greater than 6, which might indicate a blood clotting abnormality. Further review is then performed to determine whether this condition is caused by improper anticoagulation management or by disease progression.

Complete results for chart reviews by two nurse reviewers and a final reconciliation including a physician were required for inclusion in the study. The reviews began in August 2004 and ended in March 2008. The first 5 months of reviews were treated as training cases to establish a working process for ongoing reviews. At two of the three sites reviews were captured for inclusion in the study for an approximate 2-year time period between August 2005 and September 2007. For the third site, reviews were captured between August 2006 and September 2007. Analysis included over five hundred cases at each of two sites and less than two hundred at the third.

Adverse events and harm

The IHI GTT detects adverse events defined as harm to the patient as a result of medical diagnosis or therapeutics, irrespective of error or preventability. Adverse outcomes caused solely by underlying disease or by the intended consequences of treatment were not considered adverse events. Each adverse event was categorized for harm to the patient according to the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) categories [11]. The NCC MERP categories range from A (near miss, no event) to I (adverse event directly contributed to death). For the purposes of this study, we were concerned only with events categorized as E or higher. (See Appendix 2 for definitions of each level of harm.)

Data sources

The data used in this study came from reviews of medical records collected through two main sources: (1) a decision support system (DSS) used to provide demographic, diagnosis and procedure information for each patient hospitalization and (2) a specifically designed web-based data capture tool (Mayo Global Trigger Toolkit) developed not only to allow the uniform collection of data between each of the three

sites, but also to streamline the review workflow, reconciliation process and data capture of all evaluations performed by reviewers. 'Cases selected for review were identified for each 2-week period from the DSS.'

Statistical analysis

The unit of analysis for this study was the patient hospitalization. The Cohen Kappa coefficient was used as a measure of inter-rater agreement for all of three groups consisting of each of the nurse reviewers, both with each other and with a physician reviewer. Agreement between reviewers on the presence of a trigger was defined in one of two ways: (1) either both reviewers indicate the presence of the trigger event during their independent review of the patient's chart or (2) both reviewers fail to indicate the presence of the trigger event. All else was considered disagreement. Similarly, agreement on events was considered present if both reviewers determined either: (1) one or more adverse events occurred or (2) no adverse events occurred. Agreement on harm to the patient was determined at the granular level of NCC MERP categories E through I and was considered agreement when both reviewers indicated the same category of harm. In cases with multiple events per hospitalization, the highest level of harm for any of the events was utilized in the analysis.

Kappa statistics were calculated for two-level categorical data (Yes, No) for detection of 'triggers' and adverse events, and weighted Kappa, with weights for degree of disagreement for five-level data (E–I) for level of harm caused by adverse events. Similar analyses were performed overall and for each of the three institutions. It was believed that the same pairs of reviewers would increase consistency in the analysis of agreement between each of the nurse reviewers and the physician reviewer. It was possible, due to low staff turnover and high review volume per reviewer, to designate specific groups of reviewers as reviewer one and reviewer two at one site. This method was impossible in the other two sites due to budget and staffing constraints. At these sites reviewers one and two for each case were based alphabetically on reviewer identification.

Results

Out of 1138 randomly chosen cases from three distinct geographical areas of the country, 307 (27.0%) hospital stays were identified with adverse events after the final review. The percent of cases with adverse events differed significantly between institution (A: 23.1% 95% confidence interval (CI): 19.4%, 27.1%; B: 27.2% 95% CI: 23.4%, 31.4% and C: 37.9% 95% CI: 30.4%, 45.9%) but did not differ significantly by year (2005: 23.9% 95% CI: 20.2%, 27.8%; 2006: 26.7% 95% CI: 20.1%, 34.1% and 2007: 30.4% 95% CI: 26.3%, 34.8%).

Frequency of trigger identification

When comparing the triggers themselves, the frequency of individual trigger selection by reviewers varied substantially with some triggers being relatively common ('anti-emetic use' was identified by at least one reviewer on 364 (32.0%) cases), while other triggers were quite rare. 'Change in anesthetic during surgery', 'intra-op or post-op death', 'maternal/neonatal transport/transfer' and 'pathology report normal or unrelated to diagnosis' were never identified. Additionally, the yield, or the positive predictive value, of individual triggers (the rate at which a trigger was associated with an adverse event) varied. For triggers identified on at least 20 cases, the trigger with the highest yield was 'Return to surgery', where 80.6% of cases with the trigger had an adverse event, while 'X-ray intra-op or in PACU' was identified on 145 cases overall but was associated with only 37 (25.5%) cases with an adverse event. Table 1 provides the frequency and yield for the triggers with 15 or more occurrences. Of the 18 triggers seen from 1 to 14 times, pneumonia onset (11 out of 13), INR greater than 6 (7 out of 11), in-hospital stroke (3 out of 4) and removal/injury or repair of organ (3 out of 4) had positive predictive values greater than 63%.

Association between triggers, adverse events and level of harm

The average raw number of triggers identified per case was higher for those who suffered an adverse event. A total of 3361 unique triggers were identified in the 1138 cases of this study. A total of 1465 (43.6%) of these were identified in the 307 (27.0%) cases with an adverse event. More than 15% of cases with an adverse event had more than seven unique triggers identified versus less than 1% of cases with more than seven triggers among those with no adverse event. In addition, more than a quarter of cases with no adverse event had zero triggers. The mean number of uniquely identified triggers per case with at least one adverse event was 4.7 (median 4) versus cases with no adverse event having a mean of 1.8 (median 1) uniquely identified triggers ($P < 0.001$). This relationship between the presence of more triggers among cases with adverse events was observed overall for all sites collectively (Table 2), as well as in each site independently.

Of the 307 cases with adverse events, 156 (50.8%) were categorized with E level harm according to the NCC MERP Index for Categorizing Medical Errors. E level harm is classified for events that result in temporary harm that require intervention. A further 126 (41.0%) cases were categorized with F level harm (temporary harm which required hospitalization) and the remaining 25 (8.1%) cases were distributed among G, H and I levels of harm (events that result in permanent harm, require intervention to sustain life and result in death, respectively). Events with E level harm had a mean of 3.75 (median 3) uniquely identified triggers per case, whereas events with F and G or higher levels of harm had a mean 4.83 (median 4) and a mean 10.00 (median 9), respectively, further demonstrating the trend noted above of cases with greater adversity having more triggers.

Inter-rater reliability

Table 3 provides the level of agreement between nurses on the presence of any trigger and the agreement between all

Table I Frequency, positive predictive value and inter-rater agreement on individual triggers

Trigger name	All cases, $N = 1138$ (% total)	Adverse events, $N = 307$ (% yield)	Simple Kappa coefficient		
			Coefficient	Lower CI	Upper CI
Anti-emetic use	364 (32.0)	113 (31.0)	0.756	0.713	0.799
Health-care-associated infection of any kind	304 (26.7)	167 (54.9)	0.614	0.5563	0.671
Transfusion or use of blood products	274 (24.1)	123 (44.9)	0.812	0.7703	0.854
Readmission within 30 days	212 (18.6)	99 (46.7)	0.649	0.5850	0.713
Any procedure complication	164 (14.4)	107 (65.2)	0.246	0.1570	0.335
X-ray intra-op or in PACU	145 (12.7)	37 (25.5)	0.642	0.5659	0.719
Insertion of arterial or central venous line during surgery	129 (11.3)	49 (38.0)	0.742	0.6729	0.811
Care: other	122 (10.7)	61 (50.0)	0.142	0.0491	0.235
Benadryl (diphenhydramine) use	114 (10.0)	49 (43.0)	0.754	0.6827	0.825
Abrupt drop of greater than 25% in Hemoglobin or Hematocrit	100 (8.8)	58 (58.0)	0.477	0.3688	0.584
X-ray or Doppler studies for emboli	93 (8.2)	46 (49.5)	0.435	0.3216	0.548
Abrupt medication stop	91 (8.0)	47 (51.6)	0.230	0.1175	0.343
Admission to intensive care post-op	83 (7.3)	41 (49.4)	0.520	0.4055	0.634
Medication: other	62 (5.4)	39 (62.9)	−0.024	−0.0320	−0.017
Partial thromboplastin time greater than 100 s	57 (5.0)	31 (54.4)	0.503	0.3638	0.642
Pressure ulcers	57 (5.0)	35 (61.4)	0.420	0.2751	0.565
Transfer to higher level of care	56 (4.9)	37 (66.1)	0.308	0.1605	0.456
Over-sedation/hypotension	43 (3.8)	16 (37.2)	0.114	−0.0204	0.248
Vitamin K administration	37 (3.2)	19 (51.3)	0.268	0.0917	0.444
Operative time greater than 6 h	36 (3.2)	21 (58.3)	0.457	0.2783	0.636
Restraint use	33 (2.9)	23 (69.7)	0.588	0.4203	0.756
Return to surgery	31 (2.7)	25 (80.7)	0.615	0.446	0.783
Rising BUN or serum creatinine greater than two times baseline	30 (2.6)	16 (53.3)	0.323	0.122	0.523
Positive blood culture	29 (2.6)	18 (62.1)	0.542	0.355	0.730
Dialysis	28 (2.5)	13 (46.4)	0.778	0.644	0.912
Glucose less than 50 mg/dl	25 (2.2)	14 (56.0)	0.522	0.317	0.728
Patient fall	21 (1.8)	9 (42.9)	0.640	0.442	0.838
In-unit procedure	20 (1.8)	15 (75.0)	0.254	0.016	0.491
Time in ED greater than 6 h	19 (1.7)	6 (31.6)	0.475	0.233	0.717
Mechanical ventilation greater than 24 h post-op	17 (1.5)	14 (82.3)	0.449	0.189	0.709
Change in procedure	17 (1.5)	10 (58.8)	0.375	0.110	0.641
Clostridium difficile positive culture	16 (1.4)	12 (75.0)	0.311	0.040	0.583
Intubation/re-intubation	16 (1.4)	12 (75.0)	0.111	−0.100	0.322
Other[a]	125 (11.0)	59 (47.2)	NA	NA	NA

[a]Eighteen triggers were observed from 1 to 14 occurrences each, while four triggers were not observed. For example, INR < 6 was observed 14 times. CI, confidence interval; PACU, post anesthesia care unit; BUN, blood urea nitrogen; ED, emergency department.

reviewers on the presence of an adverse event, both overall and by site. Agreement between nurses was good on both triggers and adverse events with the mean Kappa ranging across sites from 0.53 to 0.73 for triggers and from 0.40 to 0.60 for adverse events. The agreement between each nurse and the physician assessment was higher ranging from 0.65 to 0.77 for the presence of an adverse event. A change in procedure occurred in one of the three institutions in May of 2007. This resulted in a switch from the use of a large group of rotating float nurse reviewers to a smaller group of more highly trained, dedicated nurse reviewers. The result of this was only a small change in the level of agreement between the engaged parties, with all Kappa results different by less than 0.05 before May 2007 and after the change.

The agreement between nurses on individual triggers varied by the type of trigger, with lower levels of agreement on more subjective assessments, like abrupt medication stop (Kappa = 0.23 95% CI: 0.12, 0.34) or over-sedation (Kappa = 0.11 95% CI: −0.02, 0.25) than more objective findings, such as INR > 6 (Kappa = 0.90 95% CI: 0.76, 1.00). Findings were consistent across years (Table 4).

Table 2 Association of an adverse event with increasing numbers of triggers per case

Number of triggers	Total cases	Cases without adverse events	Cases with adverse events	Yield of total (%)	Level E harm	Level F or higher harm
0	225	225	0	0	0	0
1	242	211	31	12.8	26	5
2	205	147	58	28.3	35	23
3	156	112	44	28.2	26	18
4	111	65	46	41.4	25	21
5	73	34	39	53.4	14	25
6	46	15	31	67.4	13	18
7	25	14	11	44.0	2	9
8 or more	55	8	47	85.4	15	32

Table 3 Agreement between nurse reviewers and between nurse and physician reviewers for identifying adverse events, by site and overall

Site	Trigger/Adverse events	Reviewer pair	Simple Kappa coefficient		
			Coefficient	Lower CI	Upper CI
Overall	Triggers	RN1–RN2	0.6312	0.5804	0.6820
Overall	Adverse events	RN1–RN2	0.5106	0.4518	0.5693
Overall	Adverse events	RN–MD	0.7109	0.6775	0.7443
A	Triggers	RN1–RN2	0.7290	0.6488	0.8092
A	Adverse events	RN1–RN2	0.5965	0.5128	0.6803
A	Adverse events	RN–MD	0.7544	0.7071	0.8016
B	Triggers	RN1–RN2	0.5777	0.5034	0.6520
B	Adverse events	RN1–RN2	0.4616	0.3686	0.5545
B	Adverse events	RN–MD	0.6628	0.6058	0.7198
C	Triggers	RN1–RN2	0.5291	0.3886	0.6696
C	Adverse events	RN1–RN2	0.4026	0.2469	0.5584
C	Adverse events	RN–MD	0.7015	0.6206	0.7824

Table 4 Agreement between nurse reviewers and between nurse and physician reviewers for identifying adverse events, by year

Year	Trigger/Adverse events	Reviewer pair	Simple Kappa coefficient		
			Coefficient	Lower CI	Upper CI
2005	Triggers	RN1–RN2	0.6085	0.4583	0.7587
2005	Adverse events	RN1–RN2	0.6108	0.4704	0.7513
2005	Adverse events	RN–MD	0.6990	0.6102	0.7879
2006	Triggers	RN1–RN2	0.6314	0.5573	0.7054
2006	Adverse events	RN1–RN2	0.5262	0.4365	0.6160
2006	Adverse events	RN–MD	0.7192	0.6676	0.7707
2007	Triggers	RN1–RN2	0.6328	0.5532	0.7125
2007	Adverse events	RN1–RN2	0.4600	0.3682	0.5517
2007	Adverse events	RN–MD	0.7057	0.6552	0.7562

In the subset of cases where both reviewers of each pair of reviewers agreed that there was an event and that there was some at least temporary harm to the patient (NCC MERP Category E harm), agreement on the level of harm to the patient varied. Harm was grouped into three categories as noted above: E, F, G and higher harm. Between nurse reviewers the agreement was low across sites with the Kappa ranging from 0.26 (95%CI: −0.09, 0.62, $N = 26$) to

0.42 (95% CI: 0.22, 0.61, $N = 56$), but between nurse and physician reviewer agreement was higher, the Kappa ranging from 0.48 (95% CI: 0.33, 0.64, $N = 95$) to 0.76 (95% CI: 0.61, 0.90, $N = 72$).

Discussion

The GTT was originally developed to provide an easy-to-use method for accurately identifying adverse events (harm) and measuring the rate of adverse events over time [5]. In our experience with the GTT across three geographically disperse locations for up to a 3-year time frame, good levels of reliability were obtained at the case-level between independent nurse reviewers for both the occurrence of any trigger and the identification of an adverse event. The agreement between nurses and a physician reviewer was very good (Kappa = 0.70 95% CI: 0.66, 0.76); however, the physician review was based on the nurse's findings rather than an independent review of the medical record. This level of agreement seen in practice at the three sites over 3 years was actually higher than that reported by reviewers involved with the IHI using records for standardized training, where the Kappa statistics between primary reviewers and physician viewers ranged from 0.35 to 0.60 on their testing records [9]. Interestingly, the good level of agreement in our study persisted even at the site when the reviews were done by a large group of rotating float nurses. With a change in the process to a smaller dedicated group of nurse reviewers, there was only a slight change in the level of agreement between the two groups.

Nurse reviewer agreement for individual triggers was much more varied with the Kappa coefficient ranging from a high of 0.90 (95% CI: 0.76, 1.00) for the INR > 6 to a low of −0.02 (95% CI: −0.03, − 0.02) for other medication issues. The latter is a generic write-in field allowing nurses to add anything else they may have found related to medications. Higher agreement appears to occur among triggers that are objective and consistently recorded in selected portions of the medical record, such as laboratory values, medications given by nurses on the floor and blood products. While triggers with lower agreement tend to be more subjective (over-sedation or procedure complication), harder to detect within a time-limited review (abrupt medication stop or readmission to the emergency department) or recorded in different locations (use of epinephrine/norepinephrine during surgery found in the anesthesia record rather than the medication administration record).

Individual triggers also varied on their yield to detect adverse events. Some triggers were associated with adverse events over 70% of the time (mechanical ventilation over 24 h), while other triggers, although frequent, were not associated with the occurrence of adverse events (use of anti-emetics). The performance of individual triggers is not a focus of the GTT. In fact, due to the likelihood of multiple triggers on cases with adverse events, the GTT may not be affected by the variability of detecting individual triggers. However, to be effectively used as a screen to detect adverse events in an automated fashion, electronic triggers should have high yields with relatively few false positives.

There have been a few reports of use of portions of the total IHI GTT [8, 12] or other adverse event screening efforts [13] Using a computer screening of free text discharge summaries to look for 'trigger words' representing adverse events, Murff *et al.* [14] found almost 60% of discharges included trigger words, but after a review of medical records, only 45% of trigger words indicated an event. A similar tool used by Brennan *et al.* [15] resulted in a positive predictive value of 21% after the physician reviewer. These tools were able to increase the number of potential problems identified over voluntary reporting; however, too many false positives make currently published e-tools infeasible for routine screening.

As evidence of the latter, Bates *et al.* [13] published an evaluation at one hospital over a 4-month period of the 18 criteria which have been used for screening medical records in the Harvard studies to assess the occurrence of adverse events in hospitals. They found that positive predictive values ranged from 15% for hospital readmission to 78% for treatment criteria because of damage subsequent to an invasive procedure. Furthermore, they were able to improve specificity by evaluating combinations of criteria. Szekendi *et al.* [16] reviewed 493 triggers among 327 hospital discharges in a 3-month period at one hospital and found that the percent of triggers with adverse events identified upon review ranged from less than 10% for four triggers to over 95% for elevated INR values. This study also identified the variable positive predictive value of individual triggers for predicting adverse events.

The GTT appears to identify more events than some other methods. Among 235 randomly selected hospital patients reviewed with the GTT, while 65 (27.7%) discharges had an event with the GTT, only 11 had provider-reported events with harm and three had AHRQ PSI. Additionally, only three discharges had provider-reported events or PSI not found by GTT [7]. A further refinement of the tool may produce a higher yield of adverse event identification and possibly save time in performing the review by allowing it to become more focused. The current methodology involves a random review of completed charts to identify any of the 55 triggers. This study demonstrates that certain triggers are more closely associated with adverse events than others. Rather than a random review of charts, a more focused review on charts known to contain the triggers with high yield (outcomes) could provide greater insight into potential problems with the delivery of care (processes). Developing an automated way to identify triggers combined with chart review would facilitate not only the identification of adverse events, but also permit nurse and physician reviewers to focus on the potential cause of the event rather than the identification of the event itself. Approximately 13% of cases with one identified trigger had an adverse event; the percentage of cases with adverse events increased to 28% with two identified triggers. In general, the more the identified triggers, the greater the number and harm of adverse events. This additional finding indicating a relationship between the

presence of more triggers among cases with adverse events requires further research to determine if certain combinations of triggers are more likely to identify an event.

Our study has several potential limitations. The study was performed only at academic hospitals. Generalizations from our study may not apply to community hospitals or other institutions. However, the study was performed by independent teams of reviewers at three hospitals in the Midwest, Southeast and Southwest parts of the USA. Although our medical record review process used two nurses, each case was only reviewed by one physician. Other physicians may have disagreed with the conclusion that an adverse event occurred on a specific case. All nurses and physician reviewers did attend consistent training. It was of note that although the three institutions had significantly different rates of adverse events, they had similar yields of adverse events for triggers and similar inter-rater agreement rates. Third, although we had cases reviewed over 3 years and three sites, sample sizes were insufficient to detect changes in types of adverse events. Fourth, although the GTT process detected a high rate of adverse events, the process required a high level of reviewer resource. Owing to the expense involved, we were limited from further collection. Finally, our study is also subject to the potential bias associated with retrospective chart review.

The GTT process shows promise of consistently identifying many adverse events that may currently go unreported. However, the current process is highly resource intensive and the identification of individual triggers appears inconsistent. With more information captured in electronic records and with bigger, more powerful computers, the time appears ripe for incorporating a trigger process into the hospital electronic environment. However, administrators, physicians and information technology professionals should be sobered by the low yields observed for many of the triggers. Too many false positives in automated systems will result in unnecessary distractions, extra costs and work-arounds to override what providers may believe are superfluous warnings.

Conclusions

With adequate training, standardized processes and collaboration between reviewers, it appears using a trigger methodology can provide a fairly reliable assessment on the occurrence of adverse events among hospitalized patients. However, we found it to be very resource intensive. Automating the process of identifying individual triggers would allow a greater number of records to be screened for possible adverse events. Focusing review on triggers more predictive of an adverse event would be a better use of resources and has a greater potential of improving care delivery and patient safety.

Acknowledgements

We would like to thank all the nurses and physicians who reviewed the medical records and used the GTT over the 3 years at all the practice locations. We also would like to acknowledge Sara Hobbs Kohrt for the preparation and management of the manuscript.

Appendix 1. Global trigger tool worksheet

Cares module triggers	+	Event description and severity E–I	Medication module triggers	+	Event description and severity E–I
Transfusion or use of blood products			C. difficile positive		
Any code or arrest			PTT > 100 s		
Dialysis			INR > 6		
Positive blood culture			Glucose < 50 Mg/dl		
X-ray or doppler studies for emboli			Rising BUN/S.Creat > 2X base		
Abrupt drop in Hct > 4% or Hg > gms			Vitamins K administration		
Patient fall			Benadryl (*Diphenhydramine*) use		
Decubiti			Romazicon (*Flumazenil*) Use		
Readmission within 30 days			Narcan (*Naloxone*) use		
Restraint use			Antiemetic use		
Infection of any kind			Over sedation/hypotension		
In hospital stroke			Abrupt medication stop		
Transfer to higher level of care			Other		
Any procedure complication					
Other			ICU module triggers		
			Pneumonia onset		

(*continued*)

Continued

Cares module triggers	+	Event description and severity E–I	Medication module triggers	+	Event description and severity E–I
Surgical module triggers			Readmission to ICU		
Return to surgery			In unit procedure		
Change in procedure			Intubation/reintubation		
Admission to ICU post op					
Intubation/Reintub/BiPap in PACU			OB module		
X-ray intra-op or in PACU			Apgar < 7 at 5 min		
Intra or post-op death			Maternal/neonatal transport/transfer		
Mech Vent > 24 h posst op			Mg sulfate or terbutaline use		
Intra-op epi or nor epi use			Infant serum glucose <50		
Post-op Troponeim level > 5			3rd or degree lacerations		
Change anesthetic during surgery			Induction of delivery		
Consult requested in PACU					
Path report normal or unrelated to dx			ER module		
Insertion of art or CVP during surgery			Readmission to ED within 48 h		
Operative time > 6 h			Time in ED > 6 h		
Removal/Injury or repair of organ					
Patient Identifier ————		Total Events ————	Descriptions of the events in greater detail		
Total LOS ————					

Appendix 2. MERP harm classification

- Category E: temporary harm to the patient requiring intervention;
- Category F: temporary harm to the patient requiring or prolonging hospitalization;
- Category G: permanent patient harm;
- Category H: harm which required intervention to sustain life;
- Category I: harm contributed to death.

References

1. Wachter RM, Foster NE, Dudley RA. Medicare's decision to withhold payment for hospital errors: the devil is in the det. *Jt Comm J Qual Patient Saf* 2008;**34**:116–23.
2. Kizer K, Stegun M. Serious reportable adverse events in health care. *Adv Patient Saf* 2002;**4**:339–52.
3. Layde PM *et al.* Patient safety efforts should focus on medical injuries. *JAMA* 2002;**287**:1993–7.
4. Kohn LT, Corrigan JM, MS D. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999.
5. Institute for Healthcare Improvement, Progress in 5 million lives campaigns. Available at: www.ihi.org/IHI/Programs/campaign/campaign.htm?TabId=3. Accessed May 19, 2010.
6. Classen DC *et al.* Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991;**266**:2847–51.
7. Naessens JM *et al.* A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;**21**:301–7.
8. Resar RK *et al.* A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf* 2006;**32**:585–90.
9. Classen D *et al.* Development and evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008;**4**:169–77.
10. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events, 2nd edn. IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement, 2009.
11. Hartwig SC, Denger SD, Schneider PJ. Severity-indexed, incident report-based medication error-reporting program. *Am J Hosp Pharm* 1991;**48**:2611–6.
12. Sharek PJ, Classen D. The incidence of adverse events and medical error in pediatrics. *Pediatr Clin North Am* 2006;**53**:1067–77.

13. Bates DW *et al.* Evaluation of screening criteria for adverse events in medical patients. *Med Care* 1995;**33**: 452–62.

14. Murff HJ *et al.* Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc* 2003;**10**:339–50.

15. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. *Med Care* 1989;**27**:1148–58.

16. Szekendi MK *et al.* Active surveillance using electronic triggers to detect adverse events in hospitalized patients. *Qual Saf Health Care* 2006;**15**:184–90.