# Study of a multisite prospective adverse event surveillance system

Alan J Forster [ID] ,[1,2] Allen Huang,[3] Todd C Lee,[4,5] Alison Jennings [ID] ,[1] Omer Choudhri,[6] Chantal Backman [ID] [1,7]

Check for updates

## ABSTRACT

**Background** We have designed a prospective adverse event (AE) surveillance method. We performed this study to evaluate this method's performance in several hospitals simultaneously.

**Objectives** To compare AE rates obtained by prospective AE surveillance in different hospitals and to evaluate measurement factors explaining observed variation.

**Methods** We conducted a multicentre prospective observational study. Prospective AE surveillance was implemented for 8 weeks on the general medicine wards of five hospitals. To determine if population factors may have influenced results, we performed mixed-effects logistic regression. To determine if surveillance factors may have influenced results, we reassigned observers to different hospitals midway through surveillance period and reallocated a random sample of events to different expert review teams.

**Results** During 3560 patient days of observation of 1159 patient encounters, we identified 356 AEs (AE risk per encounter=22%). AE risk varied between hospitals ranging from 9.9% of encounters in Hospital D to 35.8% of encounters in Hospital A. AE types and severity were similar between hospitals—the most common types were related to clinical procedures (45%), hospital-acquired infections (21%) and medications (19%). Adjusting for age and comorbid status, we observed an association between hospital and AE risk. We observed variation in observer behaviour and moderate agreement between clinical reviewers, which could have influenced the observed rate difference.

**Conclusion** This study demonstrated that it is possible to implement prospective surveillance in different settings. Such surveillance appears to be better suited to evaluating hospital safety concerns within rather than between hospitals as we could not definitively rule out whether the observed variation in AE risk was due to population or surveillance factors.

## INTRODUCTION

Improving patient safety requires the minimisation of treatment-related harm. Treatment-related harm is typically measured as the sum total of: (1) adverse events (AEs) (harms caused by medical care) including preventable AEs (harm caused by errors) and (2) potential AEs (errors with the potential for harm). Numerous studies have demonstrated a high incidence of AEs and preventable AEs in hospitalised patients.[1–8] These studies have prompted significant investments to improve patient safety, which to a large extent have been unsuccessful.[9 10] The inability of hospitals to methodically measure harm has been proposed as one fundamental reason for this lack of progress.[11]

Prospective AE surveillance is one approach to methodical measurement that has the potential to overcome the well-documented limitations of other methods of AE detection.[12–24] In this method, patients and providers are observed by a trained observer to detect specific outcomes or processes (collectively called triggers).[25–28] Triggers are identified in real time and when identified, information describing the event is collected and passed on to designated experts whose role it is to determine whether the event represents an AE, a preventable AE or a potential AE (collectively termed a harm event). Prospective AE surveillance has been evaluated in different clinical setting and has been shown to be feasible[25 29–31] and acceptable to providers and decision-makers.[32] The method has been shown to be more efficient and accurate than incident reports and chart reviews.[14 30] Most importantly, it provides rich details about the events allowing for timely identification and assessment of cases which can be used to prioritise opportunities for improvement.[33 34]

Despite promising results, the benefits of prospective AE surveillance do not figure highly into most hospitals' approaches to patient safety, and many hospitals continue to rely on traditional methods such as voluntary reporting, chart reviews and scanning of administrative data.[21] Further, it remains unknown as to whether it can be applied consistently in different acute healthcare institutions. This is an important consideration because observed variations in AE detection rates across institutions

might be misinterpreted as being the result of differences in safety, when in fact they are variations in the measurement approach. We designed this study to describe the types and severity of AEs identified by prospective AE surveillance in the same clinical service in different hospitals. In addition, we aimed to describe the potential variation in surveillance programme performance when applied in different settings.

## METHODS

### Study design

This study was a multicentre prospective observational study. We performed prospective AE surveillance simultaneously and independently for 8 weeks in five acute care hospitals to determine the harm rate among the general medicine population.

### Setting and participants

This study took place on the general medicine wards at five hospitals. Hospitals A, B, C and E were academic hospitals offering tertiary and quaternary services including a level 1 trauma centre. Hospital D was a large urban community hospital offering primary and secondary care services. Hospitals A, B and D were located in Ontario, Canada; while Hospitals C and E were located in Quebec, Canada. During the study period, we monitored all patients admitted to general medical wards until they were discharged or the study concluded. The prospective AE surveillance was performed between February and April 2012.

### Data collection and outcomes

#### Description of the prospective AE surveillance system

We conducted prospective AE surveillance concurrently in each hospital. The activities associated with prospective AE surveillance are described below and elsewhere[29] but briefly include establishment of surveillance parameters, case identification and event classification.

#### Establishment of surveillance parameters

We used a list of triggers previously developed and described elsewhere.[29] We vetted the list with all site leads who, in consultation with their staff, approved it. The list includes prespecified triggers, such as abnormal laboratory results, delays in therapy, and medication administration (online supplementary appendix A).

#### Case identification

A clinical observer (hereafter referred to as 'observer') at each site identified cases. The lead at each hospital recruited an observer for their hospital (ie, one observer per hospital except at the community hospital where there were two observers who switched midway through the surveillance period). Four of the observers were registered nurses with a range of experience on medicine wards (3–12 years) (Hospitals C–E) and two were foreign trained and licensed doctors who had not obtained their Canadian licensing requirements (Hospitals A and B). Standardised training occurred at each site over a 2-week period and consisted of a presentation, familiarisation with triggers, service specific integration and hands on observation and entry of cases into a secure online data management tool called the Patient Safety Learning System (PSLS), Datix (Datix Ltd., Swan Court, London, UK).

Immediately following the training, observers independently completed 8 weeks of surveillance with a change in observers or a change in site after 4 weeks (figure 1). This switch was designed to evaluate the impact of the observer on surveillance performance. Active surveillance took place Monday–Friday from approximately 08:00 to 16:00 hours. Observers monitored and captured standard baseline information on all patients when they were admitted to the general medicine wards during the study period. All patients on the ward were continually monitored for the presence of the prespecified triggers from the time of their admission until their discharge or the study concluded.



**Figure 1** Surveillance periods.

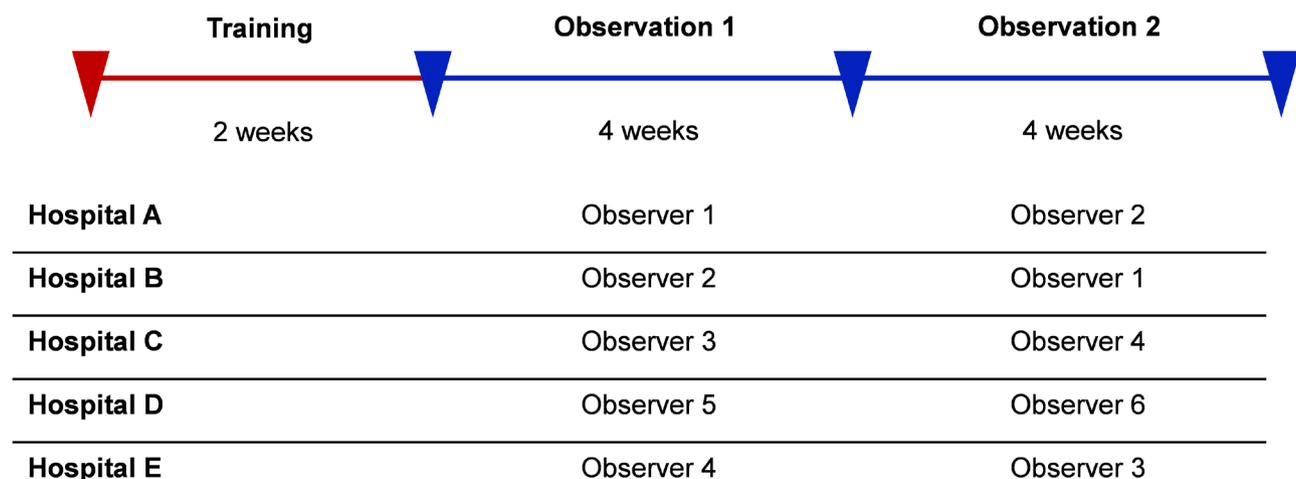| | Training | Observation 1 | Observation 2 |
|---|---|---|---|
| | 2 weeks | 4 weeks | 4 weeks |
| **Hospital A** | | Observer 1 | Observer 2 |
| **Hospital B** | | Observer 2 | Observer 1 |
| **Hospital C** | | Observer 3 | Observer 4 |
| **Hospital D** | | Observer 5 | Observer 6 |
| **Hospital E** | | Observer 4 | Observer 3 |

The surveillance activities varied slightly at each hospital, but typically consisted of obtaining the daily ward census, attending shift change reports and rounds, liaising with the nurse managers to obtain updates and incident reports, consulting nursing reports or unit log books, communicating with staff regarding specific events (ie, they could ask front-line staff questions about the case), reading discharge summaries and checking hospital information systems for abnormal lab results. These activities also allowed observers to identify events that may have occurred when they were not present on the ward. When a trigger was identified, the observer captured standard information describing the event in the PSLS.

Event classification

Once a week during the study period, a clinical review team met to review the triggers from the week. The team varied slightly at each site but typically minimally consisted of the clinical observer, a trained physician clinical reviewer and the nurse manager(s). During the meeting, the team reviewed the information entered in the PSLS for each trigger. Through discussion, consensus would be reached regarding key questions for each trigger. The questions were based on those used in the Harvard Medical Practice Study among other patient safety studies.[1–3] A six-point Likert scale was used with cut points of three to determine if an event was judged to be a potential AE, an actual AE and if it was preventable or not (where a rating of 4–6 was rated as an event, online supplementary appendix B). Responses to the questions were entered directly in the PSLS during the review and submitted for further classification.

After the clinical review, all AEs and potential AEs were classified by a single trained physician for type of event and severity. The classification for type of event was based on a modified version of the WHO International Classification for Patient Safety standards.[35] Events were classified as one or more of the following types: clinical administration, clinical process/procedure, documentation, equipment/product/medical device, patient fall, healthcare-associated infection, medication/intravenous fluid/biological treatment (includes vaccines) or nutrition. For severity, events were ranked according to the following levels of harm: nil, physiological abnormalities, symptoms, transient disability, permanent disability or death.

Analysis

We used SAS V.9.2 for all data management and analyses. We described patient baseline characteristics by calculating median and IQR for continuous variables and by using a frequency distribution for categorical variables. For disease burden, we calculated the Elixhauser index.[36] We calculated the rate of events in terms of events per 100 patient days of observation and the risk of experiencing at least one event per hospital encounter. We described events in terms of preventability, severity and type for each of the five hospitals. These measures were also broken down by observer to describe observer characteristics. To measure the rate of clinical reviewer agreement, we randomly selected 10 cases from each site (total of n=50 cases) and had the primary clinical reviewer at each hospital rate the cases (ie, reviewers rated the same cases). The proportion of cases for which the rating of harm (AEs and preventable AEs) was in agreement was measured. We assessed inter-rater reliability of reviewers using the Free-marginal Kappa statistic.

We assessed for the possible influence of (1) patient characteristics on the AE risk across sites—this was done by assessing the relationship between disease burden and AE risk; (2) observer behaviour on trigger and harm rates—this was done by comparing observer-specific and site-specific trigger and harm rates; and (3) reviewers' predilection for rating observed events as harm events (AEs and preventable AEs)—this was done by comparing the proportion of cases that were rated as harm positive (defined as the total number of harm positive cases, ie, not case specific, divided by the number of case reviews) for each reviewer.

Finally, we assessed whether there was an association of AE risk with patient, hospital and surveillance factors by performing a mixed-effects logistic regression analysis. In our model, AE risk was the dependent variable; with 'observer' being a random-effect independent variable and hospital, age, gender and Elixhauser index being fixed-effect independent variables. We repeated this analyses for preventable AEs.

## RESULTS

We observed a total 1159 patient encounters on the five general medicine wards with the patient population distributed as follows: Hospital A (n=246), Hospital B (n=235), Hospital C (n=243), Hospital D (n=313) and Hospital E (n=122).

### Patient and AE characteristics

The table 1 describes the characteristics of the patient populations at each of the participating sites. In general, patients were older adults (median age 74, IQR 61–84) which was similar across hospitals except for Hospital B whose patients were slightly younger (median age 67, IQR 56–82). There was a relatively equal gender mix across hospitals except at Hospital D where there were more females (60.7%) than males (39.3%). The most common admitting diagnosis was pneumonia (10.6%) at all sites. The second most common admitting diagnosis varied among sites although overall it was congestive heart failure (5.5%). There was an uneven distribution of chronic illness with patents in hospitals A and E having a greater burden of chronic illness and patients in hospital D having a lower burden.

**Table 1** Encounter-level descriptive statistics, by site (the percentages are column percentages)

| | | Hospital A | Hospital B | Hospital C | Hospital D | Hospital E | Total |
|---|---|---|---|---|---|---|---|
| | | 246 | 235 | 243 | 313 | 122 | 1159 |
| Age | Median (IQR) | | | | | | |
| | | 77 (62–85) | 67 (56–82) | 72 (58–81) | 79 (67–86) | 73 (55–82) | 74 (61–84) |
| Gender | N (%) | | | | | | |
| | F | 130 (52.8%) | 116 (49.4%) | 113 (46.5%) | 190 (60.7%) | 61 (50.0%) | 610 (52.6%) |
| | M | 116 (47.2%) | 119 (50.6%) | 130 (53.5%) | 123 (39.3%) | 61 (50.0%) | 549 (47.4%) |
| Top admitting diagnoses | N (%) | | | | | | |
| | All other | 154 (62.6%) | 147 (62.6%) | 152 (62.6%) | 197 (62.9%) | 80 (65.6%) | 730 (63.0%) |
| | Pneumonia | 22 (8.9%) | 33 (14.0%) | 34 (14.0%) | 23 (7.3%) | 11 (9.0%) | 123 (10.6%) |
| | Congestive heart failure | 11 (4.5%) | 14 (6.0%) | 15 (6.2%) | 18 (5.8%) | 6 (4.9%) | 64 (5.5%) |
| | COPD exacerbation | 17 (6.9%) | 4 (1.7%) | 5 (2.1%) | 11 (3.5%) | 4 (3.3%) | 41 (3.5%) |
| | Sepsis | 10 (4.1%) | 6 (2.6%) | 13 (5.3%) | 8 (2.6%) | 1 (0.8%) | 38 (3.3%) |
| | Cellulitis | 9 (3.7%) | 8 (3.4%) | 7 (2.9%) | 3 (1.0%) | 5 (4.1%) | 32 (2.8%) |
| | Other | 4 (1.6%) | 6 (2.6%) | 4 (1.6%) | 12 (3.8%) | 4 (3.3%) | 30 (2.6%) |
| | GI bleed | 4 (1.6%) | 4 (1.7%) | 4 (1.6%) | 9 (2.9%) | 7 (5.7%) | 28 (2.4%) |
| | Acute coronary syndrome | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 26 (8.3%) | 0 (0.0%) | 26 (2.2%) |
| | Acute kidney Injury | 6 (2.4%) | 9 (3.8%) | 6 (2.5%) | 1 (0.3%) | 3 (2.5%) | 25 (2.2%) |
| | GI bleed (upper) | 9 (3.7%) | 4 (1.7%) | 3 (1.2%) | 5 (1.6%) | 1 (0.8%) | 22 (1.9%) |
| Elixhauser score | Median (IQR) | | | | | | |
| | | 7 (2–12) | 6 (0–11) | 5 (0–10) | 4 (0–9) | 6 (0–14) | 5 (0–11) |

COPD, chronic obstructive pulmonary disease; GI, gastrointestinal.

Over the observation period, there were a total of 800 triggers identified (table 2). The most triggers were observed at Hospital A (n=241) and the fewest were observed at Hospital D (n=84). The AE risk, which is defined as the number of encounters with at least one AE over the total number of encounters observed varied between hospitals ranging from 9.9% in Hospital D to 35.8% in Hospital A. The AE risk was similar in Hospitals B (20.4%), C (25.9%) and E (22.1%). The AE rate per 100 patient days also varied between hospitals ranging from 1.3 in Hospital D to 6.7 in Hospital A. Hospital A also had the highest rate of preventable AEs (5.2 per 100 patient days).

The table 3 summarises AE classifications by type and severity. Of all 356 AEs detected, 45% were related to clinical processes or procedures. The second most common type of AEs were healthcare-associated infections (20%), followed by medication, intravenous fluid or biological AEs (19%). In terms of severity, four AEs (1.1%) resulted in, were associated with, or potentially led to death and two AEs (0.55%) led to permanent disability. Most AEs (56%) resulted in symptoms only. Generally, the distribution of type and severity of AEs was similar across sites. The table 4 contains examples of AEs by type.

**Potential sources of variation**

Patient characteristics

When we formally assessed the association of AE risk by a mixed-effect regression analysis using observer as the random-effect variable and using hospital, Elixhauser, age and gender as fixed-effect variables,

we identified that the driving factor was hospital, as Elixhauser, age and gender were not independently associated with AE risk (table 5). Using Hospital A as the comparator hospital, the independent ORs for AE occurrence in Hospitals B, C, D and E were, respectively, 0.42, 0.59, 0.18 and 0.49 (statistically significant). We repeated this analysis for preventable AE risk and found similar results (table 5).

Observer behaviour

We switched observers at week 4 of the surveillance period to determine the impact of observer behaviour on the results (figure 2). We assumed a similar case mix between observers. When we compared trigger detection rates within observer/hospital combinations, we observed large variation between observers. We also compared the probability triggers were rated as AEs and found variation, although of lower magnitude.

An important finding is that some observers were more likely to detect triggers but this effect was dampened by the subsequent clinical review process. There was a twofold variation in trigger detection rate between observers within Hospital D and Hospital E. Within these hospitals, observer 6 and 4 were, respectively, more likely to identify triggers than their counterparts, observers 5 and 3. The triggers detected by observers 6 and 4 were less likely to be considered AEs on subsequent clinical review than the triggers detected by observers 5 and 3, respectively. Observer 4 at Hospital E was particularly striking with less than 1 in five triggers detected as being an AE; while it was

**Table 2** Rates of AEs, by site

| | Hospital A | Hospital B | Hospital C | Hospital D | Hospital E | Total |
|---|---|---|---|---|---|---|
| | 246 | 235 | 243 | 313 | 122 | 1159 |
| **Triggers** | | | | | | |
| N | 241 | 152 | 177 | 84 | 146 | 800 |
| Mean/patient±SD | 0.98±1.36 | 0.65±1.10 | 0.73±1.17 | 0.27±0.62 | 1.20±1.46 | 0.69±1.16 |
| **Observations days** | | | | | | |
| Sum (mean/patient) | 1901 (7.7) | 1789 (7.6) | 2660 (10.9) | 2662 (8.5) | 1342 (11.0) | 10 354 (8.9) |
| **Events N** | | | | | | |
| AE | 127 | 62 | 92 | 35 | 40 | 356 |
| Preventable AE | 99 | 40 | 82 | 31 | 37 | 289 |
| Non-preventable AE | 28 | 22 | 10 | 4 | 3 | 67 |
| Potential AE | 37 | 30 | 26 | 9 | 39 | 141 |
| **Risk* N (%)** | | | | | | |
| AE | 88 (35.8%) | 48 (20.4%) | 63 (25.9%) | 31 (9.9%) | 27 (22.1%) | 257 (22.2%) |
| Preventable AE | 73 (29.7%) | 33 (14.0%) | 60 (24.7%) | 31 (9.9%) | 26 (21.3%) | 223 (19.2%) |
| Non-preventable AE | 25 (10.2%) | 19 (8.1%) | 9 (3.7%) | 4 (1.3%) | 3 (2.5%) | 60 (5.2%) |
| Potential AE | 33 (13.4%) | 28 (11.9%) | 23 (9.5%) | 9 (2.9%) | 29 (23.8%) | 122 (10.5%) |
| **Rate† (95% CI)** | | | | | | |
| AE | 6.7 (5.2–8.4) | 3.5 (2.5–4.5) | 3.5 (2.5– 4.6) | 1.3 (0.8–1.8) | 3.0 (1.8–4.3) | 3.4 (3.0–3.9) |
| Preventable AE | 5.2 (4.0–6.6) | 2.2 (4.5–3.1) | 3.1 (2.2–4.1) | 1.2 (0.8–1.6) | 2.8 (1.7–4.0) | 2.8 (2.4–3.2) |
| Non-preventable AE | 1.5 (0.9–2.1) | 1.2 (0.6–1.8) | 0.4 (0.1–0.6) | 0.1 (0.0–0.3) | 0.2 (0.0–0.5) | 0.6 (0.5–0.8) |
| Potential AE | 1.9 (1.2–2.7) | 1.7 (1.1–2.3) | 1.0 (0.6–1.4) | 0.3 (0.1–0.6) | 2.9 (1.8–4.1) | 1.4 (1.1–1.6) |

*Risk, number of encounters with at least one event/Total number of encounters observed×100%.
†Rate, total number of events/total number of days observed×100.
AE, adverse event.

almost one in two triggers for observer 3 at Hospital E. To assess this effect, we performed a two-level mixed effect model using observer and hospital as random variable and Elixhauser, age and gender as fixed effect variables. The observer effect was highly correlated with the hospital and therefore had a very small variation between observers (AE<0.001 SE too small to report; preventable AE=0.014 SE 0.058). The variation between hospitals was higher than the variations between observers (AE=0.355 SE 0.270; preventable AE=0.274 SE 0.234). Age, Elixhauser and gender were not significantly associated with AE or preventable AE risk which is consistent with the one-level mixed effect models.

**Reviewers' predilection for rating observed events as harm events**
To determine the impact of a reviewers' predilection for rating triggers as AEs, we had each reviewer rate 10 randomly selected events from each of the other hospitals. Across reviewers, the proportion of events rated as AE were similar: Hospital A reviewer (32%), Hospital B reviewer (26%), Hospital C reviewer (34%), Hospital D reviewer (32%) and Hospital E reviewer (36%) (online supplementary appendix C). The largest difference between the reviewers was between Hospital B (26%) and Hospital E (36%) for AEs. For preventable AEs, the largest difference was again between Hospital B (18%) and Hospital E (34%). The per cent overall agreement was 78.4% for AEs and 77.6% for preventable AEs. Inter-rater agreement using free-marginal kappa was 0.57, 95% CI 0.43 to 0.70 for AEs and 0.55, 95% CI 0.41 to 0.69 for preventable AEs implying moderate inter-rater reliability. It is notable that the reviewer for Hospital A appeared to have the same overall predilection for classifying cases as AEs as the reviewer for Hospital D since they rated the same number of cases as AEs and rated preventable AEs similarly. These hospitals had the highest and lowest AE rates, respectively.

## DISCUSSION
### Summary of findings
In this study, we successfully implemented the prospective AE surveillance system simultaneously on general

**Table 3** Type and severity of As, by site

| Number of adverse events | Hospital A n=127 | Hospital B n=62 | Hospital C n=92 | Hospital D n=35 | Hospital E n=40 | Total n=356 |
|---|---|---|---|---|---|---|
| Type (level 1one classification) | | | | | | |
| Behaviour | 1 (0.8%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (0.3%) |
| Clinical administration | 3 (2.4%) | 1 (1.6%) | 0 (0.0%) | 2 (5.7%) | 0 (0.0%) | 6 (1.7%) |
| Clinical process/procedure | 65 (51.2%) | 36 (58.1%) | 34 (37.0%) | 16 (45.7%) | 8 (20.0%) | 159 (44.7%) |
| Documentation | 0 (0.0%) | 0 (0.0%) | 3 (3.3%) | 1 (2.9%) | 0 (0.0%) | 4 (1.1%) |
| Equipment/product/medical device | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.9%) | 0 (0.0%) | 1 (0.3%) |
| Fall | 3 (2.4%) | 8 (12.9%) | 14 (15.2%) | 3 (8.6%) | 8 (20.0%) | 36 (10.1%) |
| Healthcare-associated infection | 20 (15.7%) | 2 (3.2%) | 31 (33.7%) | 9 (25.7%) | 11 (27.5%) | 73 (20.5%) |
| Laboratory | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.5%) | 1 (0.3%) |
| Medication/intravenous fluid/biological (includes vaccine) | 35 (27.6%) | 13 (21.0%) | 10 (10.9%) | 2 (5.7%) | 9 (22.5%) | 69 (19.4%) |
| Nutrition | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.9%) | 0 (0.0%) | 1 (0.3%) |
| Resources/organisational management | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.5%) | 1 (0.3%) |
| Transfusion medicine | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.5%) | 1 (0.3%) |
| Vascular access lines | 0 (0.0%) | 2 (3.2%) | 0 (0.0%) | 0 (0.0%) | 1 (2.5%) | 3 (0.8%) |
| Severity | | | | | | |
| Unknown | 2 (1.6%) | 1 (1.6%) | 2 (2.2%) | 1 (2.9%) | 3 (7.5%) | 9 (2.5%) |
| Physiological abnormalities | 15 (11.8%) | 5 (8.1%) | 34 (37.0%) | 12 (34.3%) | 11 (27.5%) | 77 (21.6%) |
| Symptoms | 75 (59.1%) | 40 (64.5%) | 50 (54.3%) | 15 (42.9%) | 20 (50.0%) | 200 (56.2%) |
| Transient disability | 34 (26.8%) | 13 (21.0%) | 5 (5.4%) | 7 (20.0%) | 5 (12.5%) | 64 (18.0%) |
| Permanent disability | 0 (0.0%) | 2 (3.2%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 2 (0.6%) |
| Death | 1 (0.8%) | 1 (1.6%) | 1 (1.1%) | 0 (0.0%) | 1 (2.5%) | 4 (1.1%) |

Only adverse events (excludes potential adverse events and non-events) are included in this count.

medicine wards in five different hospitals. We observed variation in the safety event rates across the five hospitals with one hospital having an increased risk of AEs and potential AEs. While it is possible the difference in rates was due to inherent safety differences, there are other possible explanations. The top patient safety concerns in all hospitals related to clinical procedures, hospital-acquired infections and medication related problems. During our implementation, we examined a variety of measurement factors, which might influence the variation in the measured rate, these included patient characteristics, observer behaviours and reviewer AE classification rate. Patient factors, including age and the burden of chronic disease among patients were not independently associated with AE risk. However, there was variation in observer behaviour and there was only moderate agreement among reviewers. Taken together, we cannot conclude that the observed variation in rates between sites was due to safety alone.

**Relevance of findings**

An inability to measure patient harm reliably is a major barrier to improving safety. The prospective AE surveillance method provides an additional approach to complement other methods. We have previously demonstrated its feasibility and acceptance by providers and hospital decision-makers.[32] A major strength of the approach is the use of observers who

**Table 4** Sample AEs by harm type, level 1 classification and severity

| AE type | Level 1 classification | Severity | Case summary |
|---|---|---|---|
| Preventable AE | Clinical process/procedure | Transient disability | Patient with history of intravenous drug use admitted for right groin swelling. Patient experienced delays in definitive management of an abscess as a result of poor coordination of care including imaging and surgical care. |
| Non-preventable AE | Clinical process/procedure | Symptoms | Elderly patient with multiple comorbidities admitted for pneumonia. Patient developed hypotension because of hypovolemia secondary to ongoing diuretic use in the setting of diarrhoea. |
| Preventable AE | Healthcare-associated infection | Death | Frail elderly patient admitted for congestive heart failure. Patient died in hospital due to complications of *Costridium difficile* colitis. |
| Potential AE | Medication/intravenous fluid/biological | Nil | Elderly patient did not receive medications in hospital as a result of delay in sending medications from the hospital pharmacy. |

AE, adverse event.

**Table 5** Mixed-effect logistic regression models

| Variable* | Model 1: any AE | | Model 2: preventable AE | |
|---|---|---|---|---|
| | OR | 95% CI | OR | 95% CI |
| Hospital B | 0.42 | 0.28 to 0.63 | 0.36 | 0.23 to 0.56 |
| Hospital C | 0.59 | 0.41 to 0.86 | 0.73 | 0.49 to 1.07 |
| Hospital D | 0.18 | 0.12 to 0.29 | 0.24 | 0.15 to 0.38 |
| Hospital E | 0.49 | 0.30 to 0.79 | 0.59 | 0.36 to 0.98 |
| Age | 1.01 | 1.00 to 1.01 | 1.01 | 1.00 to 1.01 |
| Elixhauser score | 1.00 | 0.98 to 1.03 | 1.01 | 0.99 to 1.03 |
| Female/male | 0.91 | 0.69 to 1.20 | 0.99 | 0.74 to 1.33 |

*Observer was a random variable.
AE, adverse event.

participate in hospital unit activities and consequently get to observe first-hand the unit's approach to safety management. This strength may in fact lead to a limitation if observers are inconsistent in their application of trigger detection approaches. In this study, we implemented the programme in multiple hospitals in different cities and jurisdictions. This demonstration of feasibility is an important consideration for health system leaders interested in evaluating safety in their network.

While feasible, it is important to understand the main limitation of this approach—that it is not possible to reliably discriminate the safety between hospitals. From this small sample, we cannot confirm whether the variation in hospital AE rates was due to patient safety factors alone or due to measurement effects. While this finding is important, it should be highlighted that the effect of observer variation and moderate agreement between reviewers also exists for other safety surveillance methods. The rich data obtained through surveillance remains a benefit over these other approaches and has been demonstrated to be well accepted by leaders and providers and can lead to effective quality improvement.
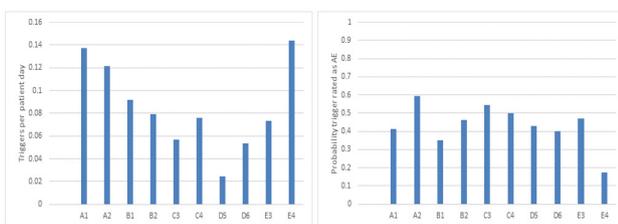


**Figure 2** Trigger rate and adverse event (AE) probability within each hospital/observer combination. Each bar represents a hospital (signified by the letter) and observer (signified by the number) combination. If observer behaviour explained the variation, then the differences within a hospital would be greater than between hospitals. Although we see some interobserver variation within hospitals, qualitative differences between hospitals persist—for example, Hospitals A and D are more different than observers A1/A2 and D5/D6. The probability triggers were classified AEs was on average 43%, with a clear outlier being observer 4 at Hospital E.

## Why AE surveillance?

Overall, the prospective AE surveillance approach has identified unit-specific patient safety problems in each of the hospitals. Previous research has demonstrated that prospective AE surveillance is more accurate in identifying patient safety incidents, compared with patient self-reporting, provider voluntary reporting or administrative methods[12–24] and that voluntary reporting and administrative data provided limited information to tailor improvement activities.[12 37 38] Prospective surveillance also provides a timely identification of AEs to allow for more rapid response to individual events.[29 32 39–41] Finally, the surveillance method of AE detection could potentially aid in the assessment of the safety culture.[41 42] The focus of future studies should aim not only detecting adverse events but also at incorporating proven methods to improve safety[32] and culture.

No prior study that we are aware of has simultaneously implemented and studied prospective surveillance in multiple hospitals within different jurisdictions and with different languages of choice. We standardised the triggers, the observer process and the physician review process. This allowed us to understand the potential benefits and limitations of a surveillance programme. This study also had some limitations. Prospective surveillance is dependent on observer behaviour and though we moved observers between facilities, we only performed one switch per hospital and we only studied five hospitals (one of which was a community hospital). While we did see an observer effect, because of the small number of institutions and observers, we need to be cautious about making conclusions—especially as there are many factors—including teaching hospital status—which could influence the results. We also had a limited ability to evaluate variability between reviewers—as we only had a small number of reviewers and they mostly performed reviews from their own hospital. This limited our ability to assess consistency of reviewers, though our demonstrations of moderate inter-rater reliability are consistent with prior studies.[43–45] To overcome these limitations, we would need to study more hospitals, observers and reviewers. Future studies could use the observer to address potential sources of variation that are attributed to the hospital in the current study, including staffing levels, overnight and weekend coverage approaches (including whether this involves house staff), clinical documentation systems and safety culture, for example. If health systems are to implement surveillance as a routine practice, then we would recommend specifically evaluating these factors, especially as these may be modifiable underlying contributors to many events.

## The business case

The business case for safety event detection is dependent on the frequency of safety events, the

likelihood of successful prevention strategies and the cost of the detection system. We have established methods for performing surveillance which are relatively inexpensive compared with the total cost burden related to adverse events. We estimate the cost per hospital for an 8-week implementation in this study to be approximately C$30 000. Studies on cost of adverse events suggest the low range of cost per event is approximately C$5000.[46] If the systematic identification of events led to interventions which reduced the annual number of cases by six, then the surveillance would be cost neutral. Of course, it may be difficult to monitor the impact of any safety strategies given the findings of this study. However, once identifying a priority safety problem using prospective surveillance, it will likely be possible to more accurately measure this specific safety concern with more precision (as objective criteria for detection can be implemented). Furthermore, this does not measure the potential for wasting efforts for poorly directed interventions that occur without data to guide them.

## Recommendations

We have several recommendations. First, we recommend using prospective surveillance as a mechanism to assess safety and set improvement priorities within a hospital or unit which has been identified as being 'at risk'. While prospective surveillance has limitations when it comes to comparing hospitals, it is highly effective at identifying safety threats at the local level and more importantly engages staff and leadership in safety assessments. Thus, it can be used to investigate and respond to units or hospitals identified using routine administrative measures such as the hospital standardised mortality ratio or composite patient safety indicators. Second, we would recommend against using the adverse event rate derived from prospective surveillance as a method to compare hospitals. For benchmarking, it is necessary to have measures with much higher reliability. To achieve this, it will be necessary to focus in on specific adverse event types, rather than the overall safety assessment used in prospective surveillance. By focusing on a specific outcome, it is possible to derive a limited number of objective criteria—as has been done for hospital-acquired infections and surgical complications. Third, if it is decided to proceed, then we would recommend the use of explicit, service-specific triggers, standard observer training and a centralised review process. While this will not remove the impact of the measurement error, it will address several of its sources. Finally, the decision to proceed to this form of surveillance should be based on numerous factors including its cost. While the programme is associated with expenses, these should be evaluated in the context of the ongoing costs of poor safety. If surveillance is used specifically in a hospital or unit with high safety threats, then it is more likely to be cost-effective.

**Author affiliations**
¹Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
²Department of Medicine, University of Ottawa Faculty of Medicine, Ottawa, Ontario, Canada
³Geriatric Medicine, The Ottawa Hospital, Ottawa, Ontario, Canada
⁴General Internal Medicine, McGill University Department of Medicine, Montréal, Québec, Canada
⁵Clinical Practice Assessment Unit, Department of Medicine, McGill University Health Centre, Montréal, Québec, Canada
⁶Internal Medicine & Critical Care, Queensway Carleton Hospital, Ottawa, Ontario, Canada
⁷Nursing, University of Ottawa Faculty of Health Sciences, Ottawa, Ontario, Canada

**ORCID iDs**
Alan J Forster http://orcid.org/0000-0003-2942-2891
Alison Jennings http://orcid.org/0000-0002-9452-3990
Chantal Backman http://orcid.org/0000-0001-7431-8159

## REFERENCES

1 Baker GR*et al*. The Canadian adverse events study: the incidence of adverse events among hospital patients in Canada. *Can Med Assoc J* 2004;170:1678–86.

2 Forster AJ*et al*. Ottawa hospital patient safety study: incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. *Canadian Medical Association Journal* 2004;170:1235–40.

3 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.

4 Thomas EJ, Studdert DM, Burstin HR, *et al*. Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care* 2000;38:261–71.

5 Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* 2001;322:517–9.

6 Wilson RM, Runciman WB, Gibberd RW, *et al*. The Quality in Australian health Care study. *Med J Aust* 1995;163:458–71.

7 Davis P, Lay-Yee R, Briant R, *et al*. Adverse events in New Zealand public hospitals I: occurrence and impact. *N Z Med J* 2002;115.

8 O'Neil AC*et al*. Physician reporting compared with medical-record review to identify adverse medical events. *Ann Intern Med* 1993;119:370–6.

9 Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* 2013;22:273–7.

10 Baines RJ, Langelaan M, de Bruijne MC, *et al*. Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. *BMJ Qual Saf* 2013;22:290–8.

11 Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* 2016;353.

12 Cullen DJ, Bates DW, Small SD, *et al*. The incident reporting system does not detect adverse drug events: a problem for quality improvement. *The Joint Commission Journal on Quality Improvement* 1995;21:541–8.

13 Beckmann U, Bohringer C, Carless R, *et al*. Evaluation of two methods for quality improvement in intensive care: facilitated incident monitoring and retrospective medical chart review*. *Critical Care Medicine* 2003;31:1006–11.

14 Flynn EA, Barker KN, Pepper GA, *et al*. Comparison of methods for detecting medication errors in 36 hospitals and skilled-nursing facilities. *Am J Health-System Pharm* 2002;59:436–46.

15 Leape LL. A systems analysis approach to medical error. *J Eval Clin Pract* 1997;3:213–22.

16 Leape LL. Reporting of adverse events. *N Engl J Med* 2002;347:1633–8.

17 Kellogg VA, Havens DS. Adverse events in acute care: an integrative literature review. *Res. Nurs. Health* 2003;26:398–408.

18 Walshe K. Adverse events in health care: issues in measurement. *Quality in Health Care* 2000;9:47–52.

19 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. ;-Apr..

20 Karson AS, Bates DW. Screening for adverse events. *J Eval Clin Pract* 1999;5:23–32.

21 Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *J Gen Intern Med* 2003;18:61–7.

22 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;127:666–74.

23 Bates DW, Evans RS, Murff H, *et al*. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;10:115–28.

24 Weinger MB, Slagle J, Jain S, *et al*. Retrospective data collection and analytical techniques for patient safety studies. *J Biomed Inform* 2003;36:106–19. ;-Apr..

25 Forster AJ, Halil RB, Tierney MG. Pharmacist surveillance of adverse drug events. *Am J Health Syst Pharm* 2004;61:1466–72.

26 Bates DW, Cullen DJ, Laird N, *et al*. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ade prevention Study Group. *JAMA* 1995;274:29–34.

27 Andrews LB, Stocking C, Krizek T, *et al*. An alternative strategy for studying adverse events in medical care. *Lancet* 1997;349:309–13.

28 Donchin Y, Gopher D, Olin M, *et al*. A look into the nature and causes of human errors in the intensive care unit. 1995. *Qual Saf Health Care* 2003;12:143–7.

29 Forster AJ, Worthington JR, Hawken S, *et al*. Using prospective clinical surveillance to identify adverse events in hospital. *BMJ Qual Saf* 2011;20:756–63.

30 Buckley MS, Erstad BL, Kopp BJ, *et al*. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit*. *Pediatric Critical Care Medicine* 2007;8:145–52.

31 Wong BM, Dyal S, Etchells EE, *et al*. Application of a trigger tool in near real time to inform quality improvement activities: a prospective study in a general medicine ward. *BMJ Qual Saf* 2015;24:272–81.

32 Backman C, Forster AJ, Vanderloo S. Barriers and success factors to the implementation of a multi-site prospective adverse event surveillance system. *Int J Qual Health Care* 2014;26:418–25.

33 Backman C, Hebert PC, Jennings A, *et al*. Implementation of a multimodal patient safety improvement program 'SafetyLEAP' in intensive care units of three large hospitals: A Cross-Case Study Analysis. *Int J Health Care Qual Assur* 2018;31:140–9.

34 Wooller KR, Backman C, Gupta S, *et al*. A pre and post intervention study to reduce unnecessary urinary catheter use on general internal medicine wards of a large academic health science center. *BMC Health Serv Res* 2018;18.

35 World Health Organization. The conceptual framework for the International classification for patient safety 2009. Jan 2. Report No.: v.1.1.

36 van Walraven C, Austin PC, Jennings A, *et al*. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical Care* 2009;47:626–33.

37 McCarthy EP, Iezzoni LI, Davis RB, *et al*. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Medical Care* 2000;38:868–76.

38 Weingart SN, Iezzoni LI, Davis RB, *et al*. Use of administrative data to find substandard care: validation of the complications screening program. *Medical Care* 2000;38:796–806.

39 Forster AJ, Fung I, Caughey S, *et al*. Adverse events detected by clinical surveillance on an obstetric service. *Obst Gynecol* 2006;108:1073–83.

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

41 Thomas EJ. The future of measuring patient safety: prospective clinical surveillance. *BMJ Qual Saf* 2015;24:244–5.

42 Stockwell DC, Kane-Gill SL. Developing a patient safety surveillance system to identify adverse events in the intensive care unit. *Crit Care Med* 2010;38(6 Suppl):S117–S125.

43 Localio AR*et al*. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. *Ann Int Med* 1996;125:457–64.

44 Thomas EJ, Lipsitz SR, Studdert DM, *et al*. The reliability of medical record review for estimating adverse event rates. *Ann Int Med* 2002;136:812–6.

45 Hofer TP, Bernstein SJ, DeMonner S, *et al*. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care* 2000;38:152–61.

46 Etchells E, Koo M, Shojania K. *The economics of patient safety in acute care: technical report*. Canadian Patient Safety Institute, 2011.