1. Title Page

DEMONSTRATION PROJECT TO REFINE, AUTOMATE AND TEST A NOVEL EMERGENCY DEPARTMENT TRIGGER TOOL

Principal Investigator and Team Members
· Richard T. Griffey, MD, MPH
· Ryan Schneider, BSN-ACNP
· Alexandre Todorov, PhD
· Aaron Papp
· Phil Asaro, MD
· Michelle Hoeschler, RN
· Tim Fortney, RN
· Christina Moran, RN
· Jamie Rolando, RN

Organization
Washington University in St. Louis School of Medicine

Inclusive Dates of the Project
09/15/2016 – 08/31/2020

Federal Project Officer
Noel Eldridge

Grant Number
R18 HS025052

Acknowledgment of Agency Support
This project was supported by grant number R18 HS025052 from the Agency for Healthcare Research and Quality.

2. Abstract

Purpose: The goals of this project were to refine, automate, and validate an Emergency Department (ED) Trigger Tool (EDTT) for adverse event (AE) detection, to allow a better estimate of the incidence of harm in the ED, to broaden the scope of event types detected, and to help direct patient quality and safety improvement efforts.

Scope: This retrospective, single-center study applied computerized queries and machine learning analytics to identify AEs in the ED from electronic medical records (EMR).

Methods: We created and validated queries to screen the EMR for 94 candidate triggers in 98,259 visits over a 13-month period, using limited Natural Language Processing (NLP) for selected triggers. We conducted two-tiered reviews to identify and categorize AEs in 5,582 records (derivation: 1,786; validation: 3,796). We identified candidate triggers associated with AEs, validating these in an independent sample. Statistical models were developed to predict AE occurrence from trigger data, enhancing EDTT yield (% record selected containing AEs) and types of AE detected.

Results: The EDTT is a promising efficient approach for detecting all-cause harm. Among 1,181 AEs, 722 were present on arrival. Twenty-nine of the 94 triggers were reliably associated with AEs. Predictive model enhanced AE yield up to 45%. This is far superior to estimates of traditional approaches used for ED AE detection while capturing a broad range of AE types. Trigger distribution data allows for population estimates of harm. NLP enhanced trigger and AE detection for selected triggers.

Key Words: adverse events, emergency department, trigger tool

3. Purpose

The emergency department (ED) accounts for 1/3 of all acute care visits and nearly 50% of hospital admissions in the US.[1-3] Yet, relatively little is known about adverse events (AEs) and the characterization of patient harm in the ED, which is often excluded from comparable in- and outpatient studies. Based on limited existing data, AEs may occur in up to 28% of ED visits,[4,5] and 71% may be preventable.[6-9] Numerous factors, such as increasing acuity, limited data for decision making, time pressures, frequent handoffs, a 24/7 work cycle, and hospital crowding and boarding, conspire to create an environment with a high potential for AEs. In many EDs, quality review consists primarily of monthly review of cases meeting blunt criteria and referrals from other departments. EDs feel compelled to review, in particular, deaths and “72-hour returns,” but these rarely involve AEs and are notoriously poor indicators of quality.[10-12] In a previous survey of eight centers, we found that only ~1.7% of records reviewed by current criteria have AEs, reflecting an inefficient process that underestimates patient harm. With surveillance methods that are porous and decades old, AEs likely go undiscovered, unreported and thus unaddressed.

Trigger tools such as those pioneered by the Institute for Healthcare Improvement (IHI) outperform traditional methods for detecting and characterizing patient harm,[13] have been developed for a number of clinical settings,[14-22] and are acknowledged as “the premier measurement strategy for patient safety.”[23,24] These consist of review of a random sample of records by a first-level reviewer (typically a nurse) for the presence of pre-defined ‘triggers’ – events that increase the likelihood an AE is present. Finding a trigger prompts an in-depth review for evidence of an AE. Putative AEs undergo confirmatory second-level review by a physician.

We recently completed a multicenter, multidisciplinary modified Delphi process to derive a first-ever ED specific Trigger Tool.[25] In this project, we refined, computerized and tested this ED Trigger Tool (EDTT). The introduction of an electronic version of a trigger tool tailored for use in the ED is a novel approach that will allow a better estimate of the incidence of harm in the ED, establish a baseline against which to measure improvement, and direct improvement efforts to provide high levels of quality and safety for ED patients. Our specific aims were:

Aim 1: To refine and test the expert consensus derived EDTT by applying a rigorous quantitative approach to evaluate 104 candidate triggers, including 47 from the consensus-derived EDTT.

We conducted two-tier manual reviews of records with dual independent first-level review and used the National Coordinating Council's Medication Event Reporting and Prevention Index to assess AE severity. We apploed classification algorithms to determine which triggers or combinations thereof were predictive of AEs and thus should be retained.

Aim 2: To create and validate an automated version of the EDTT (e-EDTT) for large-scale use with electronic medical records. We mapped triggers to electronic medical record (EMR) data fields, with sparing use of natural language processing, to automate the first step of the first-level reviews (the trigger scan). We validated the queries with manual reviews to confirm their sensitivity and specificity.

Aim 3: To validate the e-EDTT in an independent sample of records. We will scan an estimated 20,000 records using the e-EDTT to identify 6,000 records with triggers. We will review all these records, reporting the overall AE rate, event types, levels of harm, and associations with clinical and sociodemographic patient factors. We will apply rules derived in Aim 1, including combinations of triggers or factors predictive of AEs, to test whether these might further enrich the yield and efficiency of sampling.

4. Scope

4.1 Background

Emergency departments (ED) are a critical component of the U.S. healthcare system, increasingly providing primary care,[26] and accounting for over 30% of all acute care outpatient visits in the U.S.---nearly all such visits at nights and on weekends---and for over 60% of hospital admissions in the U.S.[1-3] ED visits have increased by nearly 30% between 1997 and 2007 and are on the rise.[27] In part because EDs provide medical screening and stabilization services regardless of patients' ability to pay, EDs have become a safety net for the uninsured and under-insured and for minorities.[28,29] Patients from racial and ethnic minority groups are less likely to receive routine healthcare in outpatient settings,[30-32] and such racial and ethnic disparities in healthcare may be related to a sizeable proportion of visits to urban EDs.[30-37]

Adverse events (AEs) in healthcare, defined as *the unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death,*[14] are frequent, costly, and often felt to be preventable.[38] Yet, 20 years after the IOM report "To Err is Human,"[39] and despite increased attention and resources dedicated to patient safety, AE rates have not only failed to improve but actually have worsened, raising concerns about the effectiveness of patient safety improvement efforts. AEs occur in nearly one third of hospitalizations, 1 in 7 of which result in long-term and serious harm and 44-63% of which are preventable, costing Medicare $4.4 billion annually.[40-42] There are ~440,000 deaths per year due to preventable medical errors.[43] Older estimates put AEs in the ED at up to 27.8% of visit.[5,6,8,44] However, such estimates are often for specific AE types[7] (e.g. drug events) or sub-populations (e.g., asthma, boarding patients),[8,9] with a noted lack of high-quality studies in this area.

4.2 Context

Despite its increased relevance as a source of care, relatively little is known about AEs and the characterization of patient harm in the ED. Numerous factors in the ED, such as increasing acuity, limited data for decision making, time pressure, frequent handoffs, hospital crowding and boarding, and a 24/7 work cycle, create an environment with a high potential for AEs. In many EDs, AE surveillance consists of monthly review of cases meeting blunt criteria, such as deaths and 72-hour returns, which rarely involve AEs and are notoriously poor indicators of quality.[10-12]

Table 1: AE yield by traditional criteria

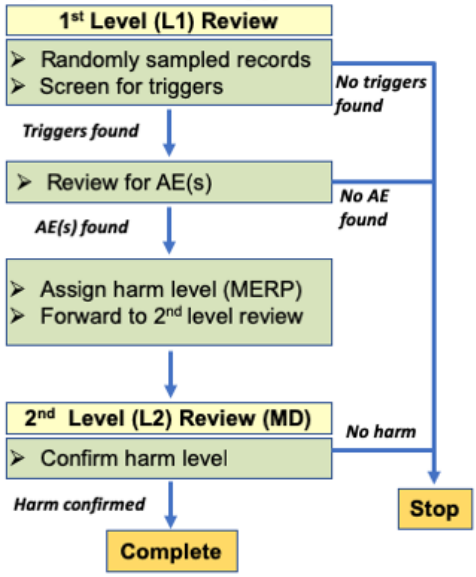
Screening Criteria	Sites using (out of 10)	Yield (%) MERP A–D (near miss)	Yield (%) MERP E–I (Harm)
Deaths	9 / 10	2.6	0.7
72hr returns	9 / 10	4.4	0.8
Upgrades in care	4 / 10	10.1	2.3
Internal referrals	5 / 10	8.9	3.0
External referrals	10 / 10	31.3	14.3
Complaints	7 / 10	1.5	0.6
	Total yield:	***6.0%***	***2.0%***

In a 2016 study, we found little variability in AE screening criteria. Referrals from other departments had the highest yield at 14.3%; all other sources were low yield (**Table 1**).[45] The overall yield in identifying AEs was under 2%, reflecting an inefficient process that underestimates harm. Typical ED quality review criteria are thus inefficient,[12] lack evidence as markers of quality,[10,46] and may miss ~90% of AEs.[13] Only 10-20% of errors are reported, 90-95% of which involved no harm to patients.[47] An unpublished American College of Emergency Physicians Quality Improvement and Patient Safety Section survey found that nearly all EDs surveyed use similar screening criteria and confirmed their low utility.[12]

An ED trigger tool could fundamentally change ED quality and safety review. Popularized by the Institute for Healthcare Improvement, and used in many countries and clinical settings,[14-22] trigger tools outperform traditional methods for characterizing patient harm, help establish a baseline to assess the results of improvement efforts over time, and guide resources for patient safety and quality improvement efforts.[13] The Mayo Clinic and AdventHealth use the IHI Global Trigger Tool (GTT) to measure and improve organizational performance. Missouri Baptist Medical Center in St. Louis decreased AEs from 35 to 9.5 per 100 admissions between 2003 and 2006, by focusing efforts on GTT-identified AEs related to hypoglycemia, VTE prophylaxis and analgesia-related over-sedation.

Acknowledged on the Agency for Healthcare Research PSNet as "the premier measurement strategy for patient safety,"[48,49] trigger tools consist of a two-stage record review (**Fig. 1**). Two first-level (L1) reviewers (typically nurses) screen randomly selected patient records for triggers (findings that make an AE more likely). If none are present, the review stops. Otherwise, in-depth reviews identify any AEs and their level of harm. Physician reviewers perform second-level (L2) review of records with putative AEs to confirm harm and severity level. Trigger tools complement incident and sentinel event reporting, claims, complaints, and concurrent chart review.[50]

Fig 1: Traditional Trigger Tool



However, the ED module of the GTT consists of just two blunt, low-yield triggers, returns within 48 hours and ED length of stay >6 hours, prompting our development of an ED-specific trigger tool (EDTT). As in prior trigger tool studies, we initially developed the EDTT using an expert consensus modified Delphi process.[25] We performed a systematic search and review for solicitation of triggers, ranking triggers based on their face validity, utility (actionability), and fidelity (sensitivity and specificity), followed by final voting on measures at an in-person meeting.[25] We then conducted a 1-year, multicenter pilot study of a 46-trigger tool at four sites (monthly random sampling of 50 records). Yield was 6.7%, outperforming traditional approaches. Ten triggers were linked to AEs. Presence of ≥1 trigger was associated with increased AE risk.[51] This encouraging pilot testing led to this demonstration project to refine and automate the EDTT.

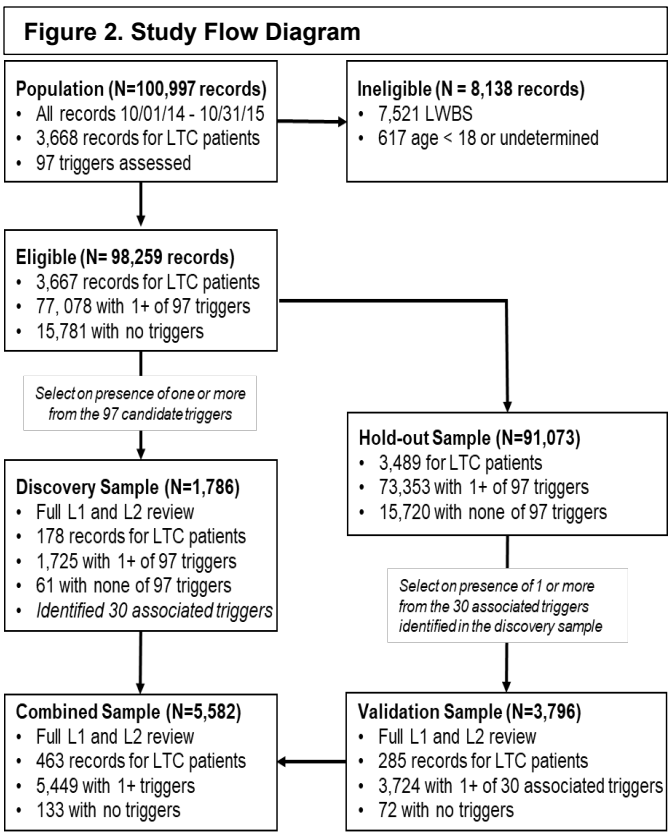
4.3 Settings
This is a retrospective observational study conducted at an urban, adult, academic medical center using data from 100,997 records between 10/1/2014 and 10/31/2015.

4.4 Participants
All ED patients aged 18 and over seen during this time frame were eligible for inclusion. Visits in which patients left without being seen, those resulting in patient elopement, or those leaving against medical advice were excluded.

5. Methods

5.1 Study design
This retrospective observational study was conducted at an urban, adult, academic medical center using data from 100,997 visits between 2014 and 2015 (**Fig. 2**). Sociodemographic data included age, race/ethnicity, and gender. Visit characteristics included date and time of arrival, Emergency Severity Index, primary diagnoses, disposition, ED length of stay, and number of providers associated with a visit. Trigger data were extracted for all records as detailed below. The extract was scrubbed to remove inconsistencies resulting from vagaries in registration and EMR systems and to verify eligibility and consistency of demographics data, to de-duplicate trigger calls and for patient de-duplication.[52] We began reviews on 12/20/2017 and concluded on 8/02/2018.

The study had two phases. During the discovery phase, we validated the computerized trigger query and performed an initial screen of each trigger for association with AEs. We then used an independent validation sample to confirm trigger-AE associations. Our work highlighted shortcomings in taxonomies for AE characterization, the importance of (i) distinguishing between events that occur in the ED and those that were present on arrival (POA); and (ii) the identification of patients from long-term care facilities (PA/LTC) as a high-risk group. We developed predictive models for both POA and ED AEs that, combined with the availability of trigger data in the entire population, allowed us to estimate ED and POA AE rates.



5.2 Trigger Query development and validation

We began with 104 of 114 candidate triggers from our Delphi process,[25] felt to be electronically capturable. We mapped 97 to structured fields in our EMR.[53] We used the AHRQ Healthcare Cost and Utilization Project Clinical Classification System to map ICD-9 codes to diagnosis-related triggers, CPT codes to procedural triggers and detailed parameters for operational, laboratory-related and other triggers. We queried home medications, chief complaint, triage data such as acuity and trauma leveling, keywords in certain free-text documentation, lab results, vital signs, documentation of certain procedures, ED diagnoses, ED disposition and orders for medications, labs, radiology studies, and consults. The query was validated against consensus manual reviews. L1 reviewers, working independently, who performed a complete trigger search in 400 records, blind to query results. Sensitivity exceeded 70% for 80 triggers. Specificity was over 92% for all. Trigger frequencies ranged from very common (e.g., C6: *ED length of stay >6 hours*, 38%) to very rare (e.g., *P7: Delivery in the ED*, 16 occurrences in ~92K visits). Most triggers had frequency <1% (62 of 97). Seven were not found at all.

5.3 NLP for Selected Triggers

We separately tested eight triggers not well captured in structured fields using natural language processing (NLP) of EMR narratives (including triage, nursing, physician, and progress notes and medical decision-making sections) with a bag-of-word (BoW) analysis on minimally pre-processed narratives. Pre-processing included recoding ED jargon, identifying common typos, tokenization, stemming, removal of stop-words, and concept grouping. We tested ~1,800 records with both EMR narratives and manual, consensus trigger ratings (randomly split 70% and 30% into training and validation sets, respectively). We identified concepts predictive of a trigger in the training set and assessed predictive performance in the validation set.

5.4 Categorizing Adverse events

AEs were rated on severity, place of occurrence (in the ED or POA—present on arrival), as acts of omission or commission, and further detailed used a novel taxonomy for ED AEs.[54] Though not part of the classic IHI approach, acts of omission were included in all our trigger tool work and in many other trigger tool studies,[55-57] because failure to diagnose and delays in, or failure to offer, treatment are significant domains of safety and quality concerns.[58] Because of the potential for subjectivity and hindsight bias as to whether omissions represent failures to act leading to AEs, we limited these to evidence-based practices and included only acts of omission with full team agreement.[59] We did not attempt assessments of preventability, felt by many to be unreliable, with no consensus on methodology and susceptible to incomplete information and hindsight and outcome biases.[60,61] We used the modified National Coordinating Council for Medication Errors Reporting and Prevention (NCC MERP) Index to rate severity, following IHI and OIG guidelines.[62] For example, level H, "*required intervention to sustain the patient's life*" is limited to interventions that occur within 1 hour.

5.5 Taxonomy development

Starting from an existing classification framework, we developed a taxonomy using an iterative process categorizing using data from 600 AEs and near misses identified during this project. We first conducted a manual search of taxonomies in use by major patient safety organizations (including AHRQ, National Patient Safety Foundation, TJC, National Quality Forum, Institute for Healthcare Improvement, WHO, Institute for Safe Medicine Practices, National Academy of Medicine, and National Coordination Council for Medication Error Reporting and Prevention). Second, a medical librarian performed a search of ontology and taxonomy databases (Linked Open Vocabularies, Basel Register of Thesauri, Ontologies, & Classifications, Taxonomy Warehouse, and BioPortal) with the phrases "adverse event" and "patient safety," and a literature review with SCOPUS 1823, using the concepts "adverse event," "adverse reaction," and "taxonomy." Our taxonomy uses a Category/Subcategory/Modifier tree for describing events, allowing up to three modifiers as needed to provide additional detail for an event. With each testing iteration, we quantified the number of times candidate categories and subcategories were used, rolling up unused or rarely used subcategories into higher-level groupings and eliminating duplicates across categories where possible. Modifiers allow us to limit the number of subcategories used while providing information that cuts across categories. For example, including service lines or particular medications as modifiers allows these to be identified as associated with events across multiple categories.

We compiled 59 scenarios encompassing the majority of potential categorizations in the taxonomy to test the reliability and accuracy of the new taxonomy. Emergency physician collaborators at three other centers and patient safety experts at Adventist Health System, working independently, each categorized these 59 AEs using the proposed taxonomy. We evaluated 1) inter-rater agreement and 2) agreement with the two authors who developed the main corpus of the taxonomy ("gold standard"). Inter-rater agreement was indexed first, by the proportion of the 59 reviews with perfect agreement (all four reviewers agree), and second, by the proportion in which the majority (3 of 4) of raters concurred, and for main category and for category + subcategory ("dyad").

5.6 Review Process

L1 reviewers completed dual independent review of each visit. L1 reviewers entered narratives for the AEs they identified and classified them as acts of omission or commission, place of occurrence (ED or POA), and severity level (NCC MERP). Evaluations extended into the first 24 hours of inpatient stays for admitted patients to look for AEs attributable to ED care. Visits in which L1s reported an AE were randomly assigned to an L2 who could agree or disagree, modify events, eliminate duplicate ones, and add missed AEs. L2s applied the taxonomy to confirmed AEs. L2s conducted in-depth reviews of ~600 records with no AEs reported on L1 review (estimated L1 false negative rate of 7.3%). As reviewers gained experience, it was very unlikely that a second L1 review detected an AE in which a first L1 review failed to do so (17 times in 2,365 such instances). The IHI GTT truncates L1 reviews at 20 minutes for much longer inpatient records. Review times were much shorter for our typically brief ED records (median: 7.4 minutes; IQR 4.7-11.9).

We set much stricter performance thresholds for reviewers than typical of other trigger tool implementations. There is a learning curve for new reviewers,[51,63] thus our emphasis on a robust training period. L1 training consisted of (1) standardized online IHI reviewer training; (2) training materials from prior IHI trigger tool studies; (3) training materials and scenarios we developed; (4) prior publications with event categorization;[59] and (5) data entry procedures. This addressed concerns regarding reliability raised in some other settings. One Swedish study found poor inter-rater reliability using the GTT,[64] but many others, including our own, reported high inter-rater reliability with sound training and monitoring procedures.[13,42,49,65,66] Our L2s performed complete records reviews until L1 sensitivity was established and continuously monitored performance with in-depth reviews of a subsample of records (L2s usually do not review the entire record, just summaries of potential AE).

Data processing, data entry, and study management were all performed via a custom web-based graphical user interface. The GUI allowed L1 and L2 data entry/editing for demographics, triggers, and AEs. The system ensured that an L1 reviewer never saw a record twice. L2 reviewers saw the two L1 reviews combined, with discrepancies highlighted. L2 reviewers could confirm, modify, or decline proposed AEs; mark one as cascading to another; add missed events; and mark an event as duplicate.

5.7 Statistical methods

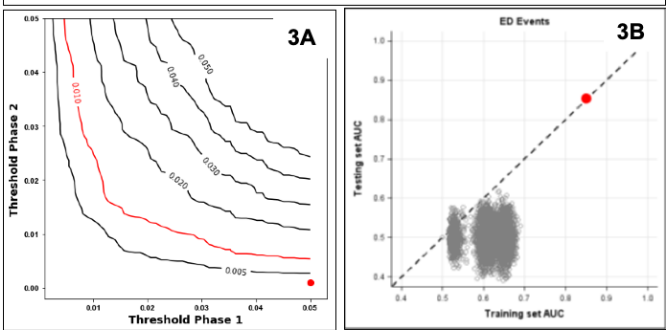
SAS 9.4 (Cary, NC) was used for database management, data preparation and descriptive and screening analyses; the Python package *Scikit-learn,*[67] for machine learning; and *R* packages, as needed. We screened for bivariate association with AEs using exact tests and with Poisson regression (AE counts as outcome, as in Landrigan et al[42]), adjusting for patient/visit characteristics, especially age and acuity.

The unadjusted analysis is required, as we did not assume that detailed visit and patient information would be available to all EDTT implementations. Generalized Estimating Equations were used to account for clustering on patient (multiple visits are common).

Machine learning (ML) was used to develop models predictive of AEs to estimate yield and AE prevalence in the population. Model performance was assessed by (1) ROC AUC (prediction ability) and Brier scores (calibration, accuracy) and for given risk score cutoffs: (2) percentage of visits meeting cutoff; (3) expected yield and percentage of captured events; and (4) expected AE types detected. Using the Scikit-learn[67] library, we fit regression-based (LASSO, ridge, elastic net), tree-based (random forests, gradient boosting) models, Bayesian neural nets, and distance-based models (SVM, K-NN). Simple LASSO logistic regression proved the more useful. Slightly higher yields could be attained with other classifiers but not enough to justify the added complexity, lower calibration degree and higher overfitting risk. The LASSO has limitations (e.g., interaction effects are not readily incorporated, and associations may be clinically obvious, but rare triggers can be *statistically* difficult to demonstrate). To avoid overfitting, we split the data into training and testing sets and used cross-validation within the training set to estimate hyper-parameters. Cross-validated calibration curves were obtained using standard methods.[68,69] The calibrated AE risk models allowed us to calculate AE risk for each of the 98K visits our population. Direct standardization[70] was then used to estimate AE prevalence.

We verified the robustness of our approach with extensive simulation studies. These showed that the risk of a single false positive for this study was well below 1% and verified that the risk of wrongly arriving at a reasonably accurate predictive model ("overfitting") using LASSO was minimal. We randomly generated AE outcomes for each of the 92,859 records (i.e., no association with any of the 92 triggers) for a 3% prevalence and selected and analyzed these "visits" exactly as in the actual study: (A) derivation: selecting on 1+ of 92 triggers and screening with Fisher's exact test with a Benjamini-Hochberg False-discovery correction (5% FDR); and (B) validation: selecting on triggers found associated in the discovery sample and again screening with Fisher's exact test (α = 0.0001). **Figure 3.A** shows the combinations of FDR and significance levels that yield the same overall risk of falsely concluding that *any* of the 92 triggers is associated with AEs, when in fact none are. Our study (red dot) is well positioned below the 1% risk curve. Our LASSO results are also well separated from those that would be expected if, in fact, there was no relationship between triggers and AEs (10,000 replicates; **3.B**).

Figure 3. Robustness analysis



6. RESULTS

6.1 Principal Findings

We reviewed a total of 5582 visits in the derivation (N=1,786) and validation (N=3,796) phases (**Table 2**). Of these 5,582 records, 5,449 had ≥1 trigger and 133 had none for comparison purposes. We detected a total of 1,181 AEs, 722 of which were POA and 459 occurred in the ED. ED AEs affected 426 visits. All but one of these 426 visits carried one or more triggers. Thus, 7.8% (425/5,449) of visits with *any* trigger had an ED AE, compared to under 1% in visits without triggers (1/133), a yield superior to traditional approaches.

Table 2: Sample Description

2A: Patient demographics

	Population N=58,497	Entire Study n=5,187	Discovery Phase n=1,755	Validation Phase n=3,554
% Female	53.7%	53.8%	53.0%	54.3%
Median age (IQR*)	42.7 (28.1—58.5)	51.5 (34.2—64.7)	51 (33.3 – 65.0)	51.9 (35.1 – 64.7)
Race				
Black	55.7%	56.5%	54.3%	58.1%
White	41.8%	41.5%	43.3%	40.1%
Other	2.5%	2.1%	2.5%	1.9%

2B: Visit Characteristics

	Population N=92,859	Entire Study n=5,582	Discovery Phase n=1,786	Validation Phase n=3,796
Acuity (ESI*)				
1- Resuscitation	2.2%	7.2%	11.1%	5.4%
2- Emergency	30.0%	40.7%	40.0%	41.1%
3- Urgent	51.8%	45.8%	43.0%	47.1%
4- Semi-Urgent	15.1%	6.1%	5.8%	6.3%
5- Non-Urgent	0.8%	0.2%	0.2%	0.3%
Disposition				
Discharged	63.7%	44.4%	42.3%	45.4%
AMA	0.7%	0.8%	0.95%	0.7%
Admit	34.2%	52.0%	52.5%	51.7%
Transfer	1.0%	1.6%	1.7%	1.5%
Expired	0.3%	1.24%	2.4%	0.7%
Other	0.03%	0.02%	0.06%	0

2C: Event Characteristics

	Entire Study 461 AEs	Discovery Phase 170 AEs	Validation Phase 291 AEs
Act of omission	48 (10.43%)	29 (17.1%)	19 (6.55%)
MERP Score*			
E	370 (80.3%)	115 (67.7%)	255 (87.6%)
F	31 (6.72%)	20 (11.8%)	11 (3.78%)
G	3 (0.65%)	0 (0%)	3 (1.03%)
H	47 (10.2%)	29 (17.1%)	18 (6.19%)
I	10 (2.17%)	6 (3.53%)	4 (1.37%)

* **IQR** = Interquartile range. **ESI** = Emergency Severity Index. **MERP** = Medical Error Reporting and Prevention index. E: resulted in the need for treatment or intervention and caused temporary patient harm; F: resulted in initial or prolonged hospitalization and caused temporary patient harm; G: resulted in permanent patient harm; H: resulted in a near-death event (e.g., anaphylaxis, cardiac arrest); I: resulted in patient death

Bivariate Association between ED AEs and individual triggers: We identified ~30 triggers as associated with AEs in the derivation phase. Twenty-four of these 30 triggers were reliably associated with AEs in the independent validation sample. Of interest, only 60% of the original consensus-derived triggers were retained; most of the remaining candidate triggers had actually been rejected during the Delphi process. This is of particular interest given that most trigger tools use consensus-derived trigger selection. Three triggers (C28, M6, and P4) that had just missed the significance threshold used in the derivation analyses demonstrated strong performance in the larger validation sample.

Table 5. Triggers associated with adverse events on bivariate analysis

	Trigger	Pop. Freq.	Discovery		Validation	
			n	PPV	*n*	PPV
L6	pCO2 >60	0.27%	34	41.2%	55	34.6%
M19	Propofol	0.81%	93	38.7%	147	36.7%
C7	BiPAP/CPAP	0.72%	38	36.8%	74	21.6%
C3	Restraint use	1.12%	114	31.6%	159	30.2%
M5	D50	0.86%	40	30.0%	103	28.2%
P8	US guided IV	1.01%	44	29.6%	120	25.0%
P6	CVC Insertion	0.53%	85	29.4%	114	32.5%
M9	Nitroglycerin/Nicardipine/Nitroprusside	1.23%	49	28.6%	126	19.8%
M3	Heparin	2.17%	61	27.9%	199	14.1%
C4	SBP<90 X2	1.51%	158	27.2%	269	29.4%
P2	Intubation	0.69%	138	26.8%	130	45.4%
C13	O2 <90%	3.45%	202	25.3%	404	21.8%
L11	BNP >300	2.66%	72	23.6%	220	14.1%
M4	Diphenhydramine	3.49%	91	23.1%	371	16.2%

C28	Chest CT for PE	1.37%	49	22.5%	90	20.0%
M17	IV Calcium	0.97%	67	22.4%	114	19.3%
M18	Opiates + Benzo	3.45%	158	22.2%	286	21.0%
C20	Temp. <35 or >38 C	2.56%	87	21.8%	215	15.8%
L2	Lactate >4	1.06%	101	21.8%	115	22.6%
M6	Peripheral pressor administration	0.29%	42	21.4%	34	44.1%
P4	Procedural sedation	0.63%	53	20.8%	78	21.8%
C8	HR>130	2.70%	147	20.4%	248	24.6%
L7	Positive troponin (>.03)	4.91%	176	19.9%	417	13.2%
C48	IV anti-hypertensives	2.95%	114	19.3%	232	16.8%
C27	SBP >180; DBP >120	9.60%	305	18.4%	702	11.7%
C19	RR >24 or <10	10.60%	453	17.9%	839	15.5%
C30	ED Boarding >6hrs	7.08%	265	15.5%	444	12.8%
L14	HCO3 <18	8.23%	306	15.0%	582	12.5%
C53	SIRS criteria	24.30%	748	13.9%	1707	11.1%

Triggers are organized in modules L-laboratories; M-medications; C-Clinical; P-Procedural; **Pop Freq.:** Frequency of trigger in entire population. ***n***: Number of times trigger seen in a particular phase; **PPV:** positive predictive value, risk of an AE if carry trigger. BiPAP – bilevel positive airway pressure; CPAP – continuous positive airway pressure; D50 – 50% Dextrose; US – ultrasound guided; CVC – central venous catheter; SBP – systolic blood pressure; O2 – oxygen; BNP - brain natriuretic peptide; CT – computed tomography; PE – pulmonary embolus; Benzo – benzodiazepine; HR – heart rate; RR – respiratory rate; SIRS – systemic inflammatory response syndrome

Multivariable modeling to predict ED and POA AEs: Twelve triggers associated with AE EDs were stably retained in multivariable analyses of all-cause harm. As noted, this analysis emphasizes model parsimony and will eliminate overlapping triggers and drop rarer triggers that do not contribute significantly to the overall AE rate. The LASSO had very good and consistent AUCs in training and testing sets (85% and 82%; 95% CI: 84%-86% and 79%-85%, respectively) and is well calibrated (Brier score = 0.06). There were no differences with results obtained via GEE, showing that repeat visits by some patients did not impact the results. We developed a LASSO model for detecting POA AEs from the trigger data alone, with good AUC (80%; 95% CI: 78%-82%), PPV (64.0%; 95% CI: 58.5% - 69.0%) and NPV (91.0%; 95% CI: 90.3% - 91.8).

Enhanced record selection approaches: Record selection based on *number* of triggers (e.g., >4 triggers) clearly enhances yield (**Table 6**) but does not leverage the differential strengths of association of individual triggers with AEs. We found that using LASSO derived risk scores consistently outperformed simple counts, for the same expected percentage of records selected. LASSO yields are 21% (CI: 18%-24%) for records in the top 10% of risk scores and as high as 49% (CI: 48%-50%) on a select 0.4% of visits. In an ED with 95,000 annual visits (~7,900/month), there would be 395 high-yield records in the top 10% of risk scores from which to select records for review. Further, the LASSO detected a larger proportion of visits than trigger counts for the same percentile selection (**Fig. 4**). Finally, selective algorithms improve efficiency and yield but in theory could narrow the range of AE types detected. Our preliminary data show no meaningful change in event type with stricter selection thresholds (**Table 7**).

Figure 4

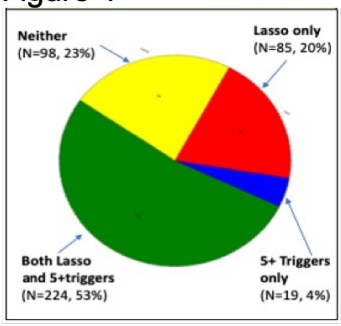


Table 6: Estimated yields comparing trigger count and weighted trigger approaches (95% CI)

# Triggers	Selection Percentile	# Visits qualifying	Expected yield using Trigger counts	Comparison with LASSO
Any	0.0%	92,859	3.0% (3.0%-3.1%)	3% (3.0%-3.1%)
1+	55.8%	41,021	5.8% (5.8%-5.9%)	6% (6.1%-6.2%)
2+	76.0%	22,297	8.2% (8.1%-8.4%)	10% (9.5%-9.7%)
3+	86.4%	12,600	10.9% (10.8%-11.1%)	13% (13.1%-13.4%)
4+	92.2%	7,265	13.8% (13.5%-14.1%)	18% (17.3%-17.7%)
5+	95.4%	4,255	16.9% (16.4%-17.3%)	22% (21.6%-22.3%)
6+	97.2%	2,566	20.5% (19.9%-21.0%)	27% (26.3%-27.4%)
7+	98.3%	1,584	24.4% (23.6%-25.2%)	32% (31.1%-32.4%)

8+	98.9%	983	28.1% (27.1%-29.2%)	38% (37.1%-38.2%)
9+	99.4%	590	33.5% (32.1%-34.7%)	44% (42.9%-44.7%)
10+	99.6%	361	38.1% (36.5%-39.9%)	49% (47.8%-50.0%)

Table 7: Distribution of event types in the entire sample and after selection

	Entire Sample (461 events)	8+ triggers (153 events)	LASSO 1% (189 events)
Device	13.5%	15.0%	12.2%
Medication	65.1%	60.1%	67.7%
Patient Care	14.5%	13.1%	10.6%
Surg./ Proc.	5.7%	10.5%	9%
HCAI	0.9%	1.3%	0.5%
Ancillary Serv.	0.4%	0%	0%
Total [2]	***100%***	***100%***	***100%***

Natural Language Processing (NLP): Using NLP and BoW analysis, we captured significant numbers of AEs not detected by the query for three concepts (Aspiration, Fall, and Delirium), less so for angioedema, reversal agents, anaphylaxis, and ataxia (**Table 8**). With *Aspiration* as example, we found ~50 over-represented words, including *aspiration* proper and derivatives (*aspirated*; typos such as *aspirate*), jargon (*NRB-non-rebreather; ETT-endotracheal tube*), and related concepts (*pneumonia*, *lactate*). Compared to the manual consensus rating, a logistic model with these words as predictors had a validation-set AUC of 98% with 11 disagreements in 1819 records. This particular NLP/BoW-derived predictor was associated with a 39% ED AE rate (CI: 19%-50%).

Table 8: # AEs captured by query and NLP (subset of 331 AEs)

Trigger	Query Alone	NLP alone	Both Query & NLP
Aspiration	2	21	12
Fall	4	51	36
Delirium	3	61	13
Angioedema	6	7	25
Reversal Agents	15	2	0
Anaphylaxis	32	5	4
Ataxia	6	4	1
Total captured	***68***	***151***	***91***

A taxonomy of ED adverse events provides better detail for quality improvement: We created a detailed three-branch taxonomy with categories, associated subcategories, and up to three modifiers to describe AEs. Better descriptions of AEs enhance the ability to identify threats and areas for intervention. For example, *Medication* is less informative than *Medication/Hypotension/Propofol, Procedural;* or *Medication/ Bleeding event/GI*. Our four raters performed well using the taxonomy. Reviewer agreement with the gold standard was 92% at the category level and 88% at the Category/Subcategory dyad level. Performance of individual raters ranged from near perfect (98%, 58/59) to very good (88%, 52/59) at the main category level and from 97% (57/59) to 78% (46/59) at the dyad level. Agreement between raters was also very good. At the main category level, three of four in four raters concurred in 55/59 scenarios (93%), and all four concurred in 46 out of 59 scenarios (78%). At the dyad level, there was perfect agreement in 40 of 59 (68%) scenarios and majority agreement in 55/59 instances (93%). There were six scenarios with no majority agreement. These were discussed in a follow-up communication and two categorizations were modified to the consensus decision.

The construct “AE” is heterogeneous, influencing which triggers are retained in models. As in other studies, the broad AE categories of *Medication* and *Patient Care* were the most common (65% and 14%, respectively). However, the ED taxonomy highlighted important differences: for example, hypotension is the most common medication-related event in the ED (36%) but is an uncommon POA event (6%). In contrast, bleeding represents 37% of POA medication-related events but only 4.3% of those in the ED. Modeling specific kinds of AEs may result in more predictive trigger sets and models that, when combined, could be higher yield for all-cause harm. Separate models for medication, patient care, device and surgical/procedural had AUCs ranging from 75% to 85%, with little overlap in the triggers retained. These models were combined using ensemble learning, for an all-cause ED AE yield of 41% (CI: 36%-46%) when selecting the top 1% of visits. These preliminary results are promising, but cited with caution, given low numbers when focusing on AE subtypes (e.g., only 62 AE device events).

Population Prevalence of Adverse Events in the ED: We used the LASSO models and direct standardization to estimate, that, in our population, 7.6% of visits have one or more POA AEs, and 4.1% of all visits incur one or more ED AEs (95% CI: 6.9%–8.2% and 3.6%–4.6%, resp.). In the 92,860 visits in our study period, we expect 7,320 POA AEs and 3800 ED AEs. This indicates that, in our population, ~66% of all AEs during that period were POA (close to the 61% in our selected study sample). POA AEs were largely acts of commission (87%), and 13% were acts of omission. POA events were more likely to result in hospitalization than ED AEs (54% vs 7%, $p < 0.001$; 95% CI: 50%-58% and 4%-9%) and less likely to result in temporary harm (30% vs 80%, $p < 0.001$; 95% CI: 27%-34% and 77%-84%).

Post-Acute and Long-Term Care (PA/LTC) Patients Account for a Disproportionately High Number of Adverse Events in the Emergency Department: In the previous analysis, we noted that patients with POA AEs tended to be older and more likely to come from a skilled nursing facility (25% vs 6% amongst those without POA event). We identified PA/LTC patients using queries of specific fields in the nursing documentation section of the EMR and natural language processing of EMR narratives. This computerized search was highly reliable compared to manual review (specificity=98%, sensitivity=91%; N=1,786 records). PA/LTC patients comprised only 8.2% of our sample but accounted for 21% of all AEs (26% POA; 13% in ED). Further, the type of POA AEs experienced by PA/LTC and non-PA/LTC patients were significantly different (**Table 9**). In PA/LTC patients, POA AEs were more likely to be patient care related (38.6% vs. 4.7%); the most common subtypes were falls (48%), traumatic injury (11%), and pressure ulcers (12.3%). In contrast, in non-PA/LTC patients, falls and traumatic injury accounted for only 8% of patient care-related AEs. In non-PA/LTC patients, POA events were more likely to be medication related (63.3% vs. 33.9%). Bleeding events were the most common subtype of medication-related events in both groups (47% and 35% of AEs). Similar trends were also present for AEs occurring in the ED. Those experienced by PA/LTC patients were more likely patient care related (25.9% vs 12.9%) and less likely to be medication related (41.4% vs 68.5). Hypotension was the most common subtype of medication-related ED AE (54.2% and 34.8% of events, for PA/LTC and non-PA/LTC, respectively).

Table 9. AEs by type and occurrence stratified by Post-Acute/Long-term care patient status

	1,179 AEs detected overall affecting 1,015 patients*			
	PA/LTC patients 247 AEs; 189 patients		Non-PA/LTC patients 932 AEs; 831 patients	
AE Type	**Present On Arrival 189 AEs 160 patients**	**ED 58 AEs 49 patients**	**Present On Arrival 529 AEs 501 patients**	**ED 403 AEs 373 patients**
Medication	64 (33.9%)	24 (41.4%)	335 (63.3%)	276 (68.5%)
Patient Care	73 (38.6%)	15 (25.9%)	25 (4.7%)	52 (12.9%)
HCAI	30 (15.9%)	1 (<1%)	68 (12.9%)	3 (<1%)
Surgery/ Procedural	8 (4.2%)	7 (12.1%)	74 (14.0%)	19 (4.7%)
Device	10 (5.3%)	11 (19.0%)	27 (5.1%)	51 (12.7%)
Care Coordination	4 (2.1%)	0 (<1%)	0 (<1%)	0 (<1%)
Ancillary Services	0 (<1%)	0 (<1%)	0 (<1%)	2 (<1%)

Conclusions

The EDTT is a promising efficient and high-yield approach for detecting all-cause harm to guide quality improvement efforts in the ED. In this single-site study of the EDTT, we observed high levels of validity in trigger selection, yield and representativeness of AEs, with yields that are superior to estimates for traditional approaches to AE detection. Record selection using a weighted trigger score outperformed a simple trigger count threshold approach and far outperformed random sampling from records with at least one trigger. We implemented a computerized query that eliminates the time-consuming manual screen for triggers and empirically evaluated a broad set of candidate triggers for associations with AEs, avoiding reliance on expert consensus alone for trigger selection. We used natural language processing (NLP) in narrative areas of the electronic medical record (EMR) to capture triggers that did not map well to structured fields. We developed a novel taxonomy of ED AEs, providing finer granularity regarding AE types detected.

We were able to derive and validate predictive models for AEs with good properties. Together with the availability of trigger data for all visits in a 1-year period, these allowed us to estimate, for that year, the prevalence of AEs in the ED (4.1% of visits) and that of AEs present of arrival (7.6% of visits). The identification of triggers and models associated with ED and POA AEs highlights the potential of the ED Trigger Tool to identify these events as part of routine surveillance reviews for quality improvement. They also shed new light on the role of the ED not only as a safety net providing acute unscheduled care for patients but also as a safety net for patient harm that occurs across the health system. Indeed, the majority (66%) of AEs detected in the ED were present on arrival. These tended to be more severe and involve older patients, with distinct patterns of event types.

Going one step further, we found that Post-Acute Care patients account for a small proportion of overall ED visits but a disproportionate number of AEs (~21% of all AEs identified). About two thirds of PA/LTC patients seen in the ED are admitted, which is double the rate for non-PA/LTC patients. ED visits for AEs among PA/LTC patients are most frequently for patient care-related events, followed by medication events and infections. Data gathered using the ED Trigger Tool could be used to provide feedback to PA/LTC facilities on either individual patients or in a summative way. These data, particularly if collected over time and applied across a health system, might also to be used help identify outlier facilities or outlier problems with quality of care.

Significance

The trigger tool approach is a novel idea for the ED. There have been no substantial changes in the overall approach to quality and safety review in the field of emergency medicine for decades. Our goal is to improve safety and quality of care for emergency patients. Creating efficient, high-yield methods for detecting and detailing harm helps identify areas for improvement, advances work toward ultimate real-time detection and prevention. We have developed and validated an automated, evidence-based ED trigger tool that is high yield for identifying a broad scope of AEs.

A computerized query is a novel approach to improve efficiency and yield of the EDTT. We have now developed the query for two EMR systems (Allscripts and Epic). The traditional trigger tool approach randomly selects records for review *then* performs a manual trigger search. A computerized query reverses the order of this process. It eliminates the time-consuming manual screening (often finding no triggers) and allows selection of records known to have triggers and thus higher AE risk. This allows reviewers to focus their reviews on detecting harm. The ability to screen an entire population of records for triggers and having a risk model relating triggers to AEs allows estimates of the level of harm in that population. When fine-tuned with larger sets of data, it may be possible to compare of rates of harm across sites over time and variation in AE types and severity.

We applied a rigorous quantitative approach not seen in prior trigger tool studies for evidence-based trigger selection and to optimize yield. Previous trigger tools are generally consensus based without empirical backing. Selecting records for review based on presence of any single EDTT trigger outperformed traditional approaches. However, even with a refined set of triggers, selecting records based on presence of any trigger had low specificity. Yields are much higher with enhanced records selection, including selecting on a nominal number of triggers (e.g., >4) and using trigger weights to calculate a 'risk score' for each record. We improved yield to over 50% by selecting records with higher risk scores, weighing this increase in specificity (fewer false positive records) against sensitivity (capturing as many AEs as possible).

We demonstrated the utility and feasibility of basic NLP. NLP and Machine Learning have the potential to make quality and safety review more efficient. We tested basic natural language processing (NLP) in screening for triggers, intentionally restricting ourselves to key word searches and qualifiers within specified free-text sections of the record. We significantly improved capture of four triggers that were inadequately captured in structured fields. We plan to expand our use of NLP, joining trigger searches with keywords and our taxonomy to automate not only the screen for triggers but also much of the first level review for AEs.

We developed a detailed taxonomy of adverse events and near misses[54] based on that of Sammer et al at AdventHealth. One principle we adopted for the taxonomy was that, for MERP E-I events, categorizations should describe the actual harm that occurred to a patient. We thus focused on harm from the patient's perspective rather than on error.[71] This taxonomy thus includes signs and symptoms, such as constipation after opioids, medication- or procedure-related hypoxia and hypotension, delirium/ confusion, various bleeding events, surgical complications and injuries from falls, etc.

For near misses, when no harm occurs, we attempt to describe the cause or nature of the near miss. The taxonomy was developed on a large set of actual ED events collected as part of this study and pilot tested in collaboration with AdventHealth and other collaborators. This pilot test yielded very good inter-rater reliability and high performance. Used with a severity scoring system, this taxonomy can be easily used and modified by emergency departments seeking to characterize harm and non-harm events in their EDs for quality improvement purposes. This presents an opportunity to have a usable, shared language when comparing quality and safety data across sites or across time. ED researchers do use various classifications for specific types of AEs, such as adverse drug events.[18-20] However, we did not find ED-focused taxonomies that categorize the wider spectrum of events encountered in that setting.

In contrast, most trigger tools use high-level categories to describe AEs (e.g., *patient care*), limiting their ability to identify specific areas for improvement. This is especially relevant, because AEs are a heterogeneous construct, requiring thoughtful approaches to trigger selection. Triggers that predict one kind of AE (e.g., *iatrogenic pneumothorax* may not be predictive of other kinds of AEs, such as *medication-related allergic reaction*). When modeling the collective outcome, "AE," one useful trigger may push another out of the model based on the prevalence of various types of AEs, that of triggers and the strengths of their association. Although we were limited by sample size, we found that combining separate models for homogeneous outcomes (e.g., device-related AEs) could result in better capture of specific events and increased overall AE yield.

Implications

In this study, we accomplished refinement, partial automation and validation of our trigger tool. We characterized and quantified the nature of AEs detected. Analytic work further describing these AEs is ongoing and multicenter testing of the EDTT is now underway. We will test the core set of triggers we identified to determine which are universal markers of ED AEs and which may be more site-specific or more useful for specific types of events. Importantly, we have gained experience using the tool in a real-world fashion and a basis for understanding what challenges may lie ahead for disseminating and implementing this tool more broadly. We will also pursue further enhancements to yield using rules combining triggers and attempt to partly automate the first level review with machine learning approaches. Our goal would be to incorporate these features in the tool for general use. Some trigger tools have been used in real-time warning systems to help prevent anticipated harm. We can imagine a similar approach in the ED to identify triggers and potential AEs but we are probably still a few big steps away from this. Our initial focus will be on streamlining and facilitating use of the tool including creation of a free online training toolkit to lower barriers to use the trigger tool.

6. List of Publications and Products

- Griffey RT, Schneider RM, Todorov AA. The Emergency Department Trigger Tool: A Novel Approach to Screening for Quality and Safety Events. Ann Emerg Med. 2020Aug;76(2):230-240.
- Griffey RT, Schneider RM, Todorov AA. Adverse Events Present on Arrival to the Emergency Department: The ED as a Dual Safety Net. Jt Comm J Qual PatientSaf. 2020 Apr;46(4):192-198.
- Griffey RT, Schneider RM, Todorov AA. The Emergency Department Trigger Tool: Validation and Testing to Optimize Yield. Acad Emerg Med. 2020. Dec;27(12):1279-1290.
- Griffey RT, Schneider RM, Todorov AA, Yaeger L, Sharp BR, Vrablik MC, Aaronson EL, Sammer C, Nelson A, Manley H, Dalton P, Adler L. Critical Review, Development, and Testing of a Taxonomy for Adverse Events and Near Misses in the Emergency Department. Acad Emerg Med. 2019 Jun;26(6):670-679.
- Griffey RT, Schneider RM, Adler L, Todorov A. Post-Acute and Long-Term Care Patients Account for a Disproportionately High Number of Adverse Events in the Emergency Department. J Am Med Dir Assoc. 2020 Aug 11:S1525-8610(20)30565-X.
- Griffey RT, Schneider RM, Sharp BR, Pothof J, Vrablik MC, Granzella N, Todorov AA, Adler L. Multicenter Test of an Emergency Department Trigger Tool for Detecting Adverse Events. J Patient Saf. 2018 Jul 18:10.1097/PTS.0000000000000516. doi: 10.1097/PTS.0000000000000516. Epub ahead of print. PMID: 30395000; PMCID: PMC6343477.

REFERENCES

1. Pitts SR, Carrier ER, Rich EC, Kellermann AL. Where Americans get acute care: increasingly, it's not at their doctor's office. Health Aff (Millwood) 2010;291620-9.
2. Morganti KG, Bauhoff S, Blanchard JC, Abir M, Iyer N. The evolving role of emergency departments in the United States. Santa Monica, California: Rand; 2013.
3. Schuur JD, Venkatesh AK. The growing role of emergency departments in hospital admissions. The New England journal of medicine 2012;367391-3.
4. Fordyce J, Blank FS, Pekow P, et al. Errors in a busy emergency department. Ann Emerg Med 2003;42324-33.
5. Stang AS, Wingert AS, Hartling L, Plint AC. Adverse events related to emergency department care: a systematic review. PLoS One 2013;8:9 e74214.
6. Calder LA, Forster A, Nelson M, et al. Adverse events among patients registered in high-acuity areas of the emergency department: a prospective cohort study. CJEM 2010;12:5 421-30.
7. Hafner JW, Jr., Belknap SM, Squillante MD, Bucheit KA. Adverse drug events in emergency department patients. Ann Emerg Med 2002;39:3 258-67.
8. Liu SW, Thomas SH, Gordon JA, Hamedani AG, Weissman JS. A Pilot Study Examining Undesirable Events Among Emergency Department-Boarded Patients Awaiting Inpatient Beds. Annals of Emergency Medicine 2009;54:3 381-5.
9. Camargo CA, Jr., Tsai CL, Sullivan AF, et al. Safety climate and medical errors in 62 US emergency departments. Ann Emerg Med 2012;60:5 555-63 e20.
10. Rising KL, Victor TW, Hollander JE, Carr BG. Patient Returns to the Emergency Department: The Time-to-return Curve. Acad Emerg Med 2014;21:8 864-71.
11. Pham JC, Kirsch TD, Hill PM, DeRuggerio K, Hoffmann B. Seventy-two-hour returns may not be a good indicator of safety in the emergency department: a national study. Acad Emerg Med 2011;18:4 390-7.
12. Griffey RT, Bohan JS. Healthcare provider complaints to the emergency department: a preliminary report on a new quality improvement instrument. Quality & Safety in Health Care 2006;15:5 344-6.
13. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011;30:4 581-9.
14. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Cambridge, Massachusetts. Institute for Healthcare Improvement; 2009. June 22, 2020: at http://www.ihi.org/resources/Pages/IHIWhitePapers/IHIGlobalTriggerToolWhitePaper.aspx
15. Suarez C, Menendez MD, Alonso J, Castano N, Alonso M, Vazquez F. Detection of adverse events in an acute geriatric hospital over a 6-year period using the Global Trigger Tool. J Am Geriatr Soc 2014;62:5 896-900.
16. Sharek PJ, Horbar JD, Mason W, et al. Adverse events in the neonatal intensive care unit: Development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics 2006;118:4 1332-40.
17. Najjar S, Hamdan M, Euwema MC, et al. The Global Trigger Tool shows that one out of seven patients suffers harm in Palestinian hospitals: challenges for launching a strategic safety plan. Int J Qual Health Care 2013;25:6 640-7.
18. Mattsson TO, Knudsen JL, Brixen K, Herrstedt J. Does adding an appended oncology module to the Global Trigger Tool increase its value? International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua 2014.
19. Kalenderian E, Walji MF, Tavares A, Ramoni RB. An adverse event trigger tool in dentistry: a new methodology for measuring harm in the dental office. J Am Dent Assoc 2013;144:7 808-14.
20. Carnevali L, Krug B, Amant F, et al. Performance of the adverse drug event trigger tool and the global trigger tool for identifying adverse drug events: experience in a Belgian hospital. Ann Pharmacother 2013;47:11 1414-9.
21. Szekendi MK, Sullivan C, Bobb A, et al. Active surveillance using electronic triggers to detect adverse events in hospitalized patients. Quality & Safety in Health Care 2006;15:3 184-90.

22. Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf 2006;32:10 585-90.
23. Sharek PJ. The Emergence of the Trigger Tool as the Premier Measurement Strategy for Patient Safety. In: AHRQ Web M & M, ed. http://webmm.ahrq.gov/perspective.aspx?perspectiveID=1202012.
24. Agency for Healthcare Research and Quality, Perspectives in Safety. AHRQ webM&M: In Conversation with...David C. Classen. Wachter RM. at http://webmm.ahrq.gov/perspective.aspx?perspectiveID=119.
25. Griffey RT, Schneider RM, Adler LM, et al. Development of an Emergency Department Trigger Tool Using a Systematic Search and Modified Delphi Process. J Patient Saf 2020;16:1 e11-e7.
26. Cunningham PJ, Clancy CM, Cohen JW, Wilets M. The use of hospital emergency departments for nonurgent health problems: a national perspective. Med Care Res Rev 1995;52:4 453-74.
27. Niska R, Bhuiya F, Xu J. National hospital ambulatory medical care survey: 2007 emergency department summary. National Health Statistics Report 2010;261-32.
28. Tang N, Stein J, Hsia R, Maselli J, Gonzales R. Trends and characteristics of US emergency department visits, 1997-2007. Journal of the American Medical Association 2010;304664-70.
29. Garcia T, Bernstein A, Bush M. Emergency department visitors and visits: Who used the emergency room in 2007? NCHS data brief, no 38. Hyattsville, MD. National Center for Health Statistics; 2010. at
30. Bliss E, Meyers D, Phillips R, Jr, Fryer G, Dovey S, Green L. Variation in participation in health care settings associated with race and ethnicity. J Gen Intern Med 2004;19:9 931-6.
31. Olson E. No room at the inn: a snapshot of an American emergency room. Stanford Law Rev 1994;46:2 449-501.
32. Tyrance P, Jr, Himmelstein D, Woolhandler S. US emergency department costs: no emergency. Am J Public Health 1996;86:11 1527-31.
33. Baker D, Stevens C, Brook R. Regular source of ambulatory care and medical care utilization by patients presenting to a public hospital emergency department. Journal of the American Medical Association 1994;271:24 1909-12.
34. Blanchard J, Haywood Y, Scott C. Racial and ethnic disparities in health: an emergency medicine perspective. Acad Emerg Med 2003;10:11 1289-93.
35. Burt C, McCaig L. Trends in hospital emergency department utilization: United States, 1992-1999. Vital Health Stat 13 2001;1501-34.
36. Lawrenson R, Leydon G, Freeman G, Fuller J, Ballard J, Ineichen B. Are we providing for ethnic diversity in accident & emergency (A&E) departments? Ethn Health 1998;3:1-2 117-23.
37. Lowe R. Does managed care gatekeeping affect African Americans' health? LDI Issue Brief 2001;6:7 1-4.
38. Flin R. Measuring safety culture in healthcare: A case for accurate diagnosis. Safety Science 2007;45:6 653-67.
39. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. Qual Saf Health Care 2006;15:3 174-8.
40. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008;17:3 216-23.
41. Adverse Events in Skilled Nursing Facilities: National Incidence Among Medicare Beneficiaries. In: General DoHHSOotI, ed.2014:1-69.
42. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med 2010;363:22 2124-34.
43. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf 2013;9:3 122-8.
44. Fordyce J, Blank FS, Pekow P, et al. Errors in a busy emergency department. Ann Emerg Med 2003;42:3 324-33.
45. Griffey RT, Schneider RM, Sharp BR, et al. Description and Yield of Current Quality and Safety Review in Selected US Academic Emergency Departments. J Patient Saf 2020;16:4 e245-e9.

46. Pham JC, Kirsch TD, Hill PM, DeRuggerio K, Hoffmann B. Seventy-two-hour returns may not be a good indicator of safety in the emergency department: a national study. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2011;18390-7.
47. IHI Global Trigger Tool for Measuring Adverse Events. Institute for Healthcare Improvement, 2014. (Accessed September 2014, at http://www.ihi.org/resources/Pages/Tools/IHIGlobalTriggerToolforMeasuringAEs.aspx.)
48. Agency for Healthcare Research and Quality, Perspectives in Safety. AHRQ Web M&M: In Conversation with...David C. Classen. Wachter RM. April 23, 2019, at https://psnet.ahrq.gov/perspectives/perspective/119/In-Conversation-With-David-C-Classen-MD-MS?q=web+mm.
49. Sharek PJ. The Emergence of the Trigger Tool as the Premier Measurement Strategy for Patient Safety. 2012. 2105, at https://psnet.ahrq.gov/perspectives/perspective/120/the-emergence-of-the-trigger-tool-as-the-premier-measurement-strategy-for-patient-safety.)
50. Adler L, Denham CR, McKeever M, et al. Global Trigger Tool: Implementation Basics. Journal of Patient Safety 2008;4245-9.
51. Griffey RT, Schneider RM, Sharp BR, et al. Multicenter Test of an Emergency Department Trigger Tool for Detecting Adverse Events. J Patient Saf 2018;epub ahead of print DOI: 10.1097/PTS.0000000000000516.
52. Campbell KM. Link King: A SAS/AF application for record linkage and unduplication. SUGI Proceedings, 020-30; 2005.
53. Griffey RT, Schneider RM, Todorov AA. The Emergency Department Trigger Tool: A Novel Approach to Screening for Quality and Safety Events. Ann Emerg Med 2019;76:2 230-40.
54. Griffey RT, Schneider RM, Todorov AA, et al. Critical Review, Development, and Testing of a Taxonomy for Adverse Events and Near Misses in the Emergency Department. Acad Emerg Med 2019;26:6 670-9.
55. Adverse Events in Skilled Nursing Facilities: National Incidence Among Medicare Beneficiaries. Washington D.C.: Department of Health & Human Services. Office of the Inspector General; 2014. at http://oig.hhs.gov/reports-and-publications/oei/a.asp#adverse_care.
56. Matlow AG, Cronin CM, Flintoft V, et al. Description of the development and validation of the Canadian Paediatric Trigger Tool. BMJ Qual Saf 2011;20:5 416-23.
57. Lindblad M, Schildmeijer K, Nilsson L, Ekstedt M, Unbeck M. Development of a trigger tool to identify adverse events and no-harm incidents that affect patients admitted to home healthcare. BMJ Qual Saf 2018;27:7 502-11.
58. Webster JS, King HB, Toomey LM, et al. Understanding Quality and Safety Problems in the Ambulatory Environment: Seeking Improvement With Promising Teamwork Tools and Strategies. In: Henriksen K, Battles JB, Keyes MA, Grady ML, eds. Advances in Patient Safety: New Directions and Alternative Approaches (Vol 3: Performance and Tools). Rockville (MD)2008.
59. Griffey RT, Schneider RM, Sharp BR, Vrablik MC, Adler L. Practical Considerations in Use of Trigger Tool Methodology in the Emergency Department. Journal of Patient Safety 2018;epub ahead of print DOI: 10.1097/PTS.0000000000000448e1-6.
60. Wears RL, Nemeth CP. Replacing hindsight with insight: toward better understanding of diagnostic failures. Ann Emerg Med 2007;49:2 206-9.
61. Papanicolas I, Figueroa JF. Preventable harm: getting the measure right. BMJ 2019;366l4611.
62. National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors. (Accessed July 8 2016, at http://www.nccmerp.org/medErrorCatIndex.html.)
63. Landrigan CP, Stockwell D, Toomey SL, et al. Performance of the Global Assessment of Pediatric Patient Safety (GAPPS) Tool. Pediatrics 2016;137:6.
64. Schildmeijer KG, Nilsson L, Arestedt K, Perk J. The assessment of adverse events in medical care; lack of consistency between experienced teams using the Global Trigger Tool'. BMJ Qual Saf 2013;22:3 271-2.

65. Sharek PJ, Parry G, Goldmann D, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. Health Serv Res 2011;46:2 654-78.
66. Classen DC, Lloyd RC, Provost L, Griffin FA, Resar R. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. Journal of Patient Safety 2008;4:3 169-77.
67. Pedregosa F. Scikit-learn: Machine Learning in Python. Journal of Machine Learning 2011;12:Oct 2825-30.
68. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med 2014;33:3 517-35.
69. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:1 128-38.
70. Choi BC, de Guia NA, Walsh P. Look before you leap: stratify before you standardize. Am J Epidemiol 1999;149:12 1087-96.
71. Sammer C, Miller S, Jones C, et al. Developing and Evaluating an Automated All-Cause Harm Trigger System. Jt Comm J Qual Patient Saf 2017;43:4 155-65.