The Relationship between Hospital Workload and Patient Safety
Final Research Report to AHRQ

Principal Investigator: Joel S. Weissman, PhD
Co- Principal Investigator: Jeffrey Rothschild, MD
Co-Investigators and Study Staff (In alphabetical order):

David W. Bates MD, MSc
Eran Bendavid, MD
Melissa Bender, MD
E. Francis Cook, PhD
JoAnn David-Kasdan, RN, MS
Scott Evans, PhD
Peter Haug, PhD
Jim Lloyd, PhD
Harvey Murff, MD, MPH
Leslie G. Selbovitz, MD
Peter Sprivulis, MBBS, PhD

Organization:
MGH Institute for Health Policy
50 Staniford Street, 9th Floor, Suite 901
Boston, MA 02114 USA

Inclusive Dates of Project:
9/01/03 – 6/29/05

Federal Project Officer: Eileen Hogan

***Acknowledgment of Agency Support*:** This work was supported by Grants 1 R01 HS12035 and RO1 HS12035-02-S1 from the Agency for Health Research and Quality, USDHHS. The views and opinions expressed in this report are the authors' and no endorsement by AHRQ is intended or implied.

File: AHRQRpt_final_20pp.doc

Structured Abstract

Purpose: Aim 1 of this study was to determine the extent to which hospital workload pressures affect the rate at which adverse events (AEs) occur among hospitalized patients. Aim 2 was to develop a computerized tool using the electronic medical record (EMR) to monitor adverse events.

Scope: Hospitals are under pressure to operate at maximum efficiency. Whether achieving such goals puts these facilities at odds with the goal of patient safety is unknown.

Methods: We collected a random sample of 24,676 patients discharged from the medical/surgical services at four hospitals during October 2000 to September 2001. We used a structured implicit review tool for final AE designation by physician reviewers. Our main analysis employed Poisson regressions, with the patient day as the unit of analysis and with control for clustering, to predict the likelihood of an AE. For Aim 2, analyses were conducted to evaluate the accuracy and cost effectiveness of Natural Language Processing and other computerized administrative data screening adverse event detection tools in comparison to manual chart review methods.

Results: For Hospital A, the busiest hospital, admissions and patients per nurse were significant at $p<.05$, and occupancy rate, discharges, and DRG-weighted census were significant at $p<.10$. Results at other hospitals varied and were mainly nonsignificant. For aim 2, the overall sensitivity of the NLP system was 24.6%, and the overall specificity was 89.47%.

Key Words: Patient safety, workload, EMR, NLP

Purpose (Objectives of Study).

The primary purpose of this study was to determine the extent to which hospital working conditions, specifically workload pressures, affect the rate at which adverse events (AEs) occur among hospitalized patients. We defined workload pressure as a state of heightened patient care activity. Workload pressure was indicated by measures of bed occupancy and patient volume, admission/discharge throughput, and the average complexity or acuity of patients' conditions. The likelihood that AEs and related medical errors occurred more often during busy times was an idea with intuitive appeal, yet it had never been tested empirically. The second purpose of the study sought to take advantage of the data on adverse events that was collected for the first part of the study. We planned to develop a computerized tool using the electronic medical record (EMR) at one of the study hospitals that we hypothesized would be a far more efficient way to monitor adverse events than the current reliance on chart review and abstraction. The supplementary aims as formally stated were 1s) to develop and implement software using natural language processing (NLP) designed to screen for adverse events in medical and surgical patients using narrative text from electronically stored discharge summaries and 2s) to evaluate the NLP approach to identifying AEs.

Scope (Background, Context, Settings, Participants, Incidence, Prevalence).

In its report, Crossing the Quality Chasm, the Institute of Medicine called for improvements in both patient safety and efficiency.[1] Though both goals are laudable, the IOM did not address the possibility that achieving one goal might be at odds with the other. The previous IOM report, To Err is Human, described the extent of the patient safety phenomenon, stressed the importance of system design in reducing human error, and brought the issue to the attention of the public.[2] These reports emerged in the midst of great upheaval in the US health system. Hospitals and other providers are under tremendous pressure to operate at maximum efficiency (lowest cost), and many may be providing more complex care in busier units with potentially fewer resources. In some areas of the country, such as Boston, the consolidation of the hospital industry combined with downsizing of bed capacity has caused hospitals to become crowded during busy times of the year, in many cases leading to the practice of diverting patients from their emergency rooms when beds or other services are not available. Whether such heightened activity harms patients is unknown.

Using administrative data and targeted chart reviews of patients at four hospitals, we examined the association between adverse events and measures of hospital workload. Workload pressure was measured daily by such indicators as occupancy rates, days on diversion, complexity of patient case mix, and other measures of activity that might lead to staff stress. Our key study question was as follows: Do adverse events in hospitals occur more often on crowded or busy days than on other days?

The remainder of this report is divided into two parts, reflecting each aim.

Aim One: THE RELATION OF PEAK WORKLOAD IN HOSPITALS TO PATIENT SAFETY

Methods (Study Design, Data Sources/Collection, Interventions, Measures, Limitations).

Study Period: We defined the study period of interest as the 12-month period October 1, 2000, to September 30, 2001. Data from electronic data sources were requested from hospitals for the 15-month between October 1, 2000, and December 31, 2001, to ensure capture of delayed adverse events related to care during the study period.

Data: Two main types of data were needed for the project in addition to standard individual patient demographic and casemix data: first, data derived from patient records concerning adverse events; second, data derived from hospital administrative systems concerning hospital workloads, including occupancy, admission & discharge rates, overall patient group acuity/complexity information and staffing indices, including budget, target and actual staffing levels, and skill mix.

Sample: Our primary sample of patients for Aim 1 included a random sample of 24,676 patients from medical and surgical units out of 58,143 eligible patients discharged from the medical and surgical services at four hospitals during October 2000 to September 2001 (Table). In order to enrich the sample that was reviewed for AEs, we employed an administrative data flag or screen.

Table E1: Sample From Four Study Hospitals – Exclusions From Administrative Data

Hospital	Admissions	Excluded	%	Available for Screening	%	Selected Study Sample
HOSPITAL A	65,158	36,910	56.6%	28,248	43.4%	8,499
HOSPITAL B	30,710	16,017	52.2%	14,693	47.8%	7,414
HOSPITAL C	13,150	6,694	50.9%	6,456	49.1%	3,866
HOSPITAL D	18,510	9,764	52.7%	8,746	47.3%	4,897
Total	**127,528**	**69,385**	**54.4%**	**58,143**	**45.6%**	24,676

The main purpose of the screen was to identify within the sample of hospital admissions a subsample that contained a large number of AEs. A good screen, by definition, should be easy to perform and interpret, should measure something directly related to the disease, should have low risk and cost, and should be highly sensitive and specific.[3] We combined screens from David Bates et al, as described in their 1995 article,[3] the Complication Screening Program (CSP) from Iezzoni et al,[4] and an early version of AHRQ's Patient Safety Indicators (PSIs), available at the time of sample selection.[5] Though some of the PSIs appear similar to the CSPs, the overlap of groups screened positive by each tool was relatively small (about 20%). Therefore, we decided to use the CSP and PSI as two independent tools.

The study sample included both screen-positive (n=6,841) and screen-negative cases (n=17,835). For purposes of analysis, we assumed that screen-negative cases did not contain AEs. This was a conservative assumption.

Dependent Variable - Detection of Adverse Events

We developed a structured implicit review tool on a computer platform on which trained RNs with clinical experience prepare case summaries for final AE designation by two physician reviewers. The methodology used in this study to discover, identify, and rate adverse events differs from prior work in several important ways. First, we predefined a set of important and most common adverse events of primary interest (Table). Our study team put together an evidence-based case definition for each AE that was guided by the literature and clinical guidelines. The case definitions for all types of infections were obtained from the CDC's NNIS (National Nosocomial Infections Surveillance System).[6] The definitions for surgical complications were informed by conversations with academic surgeons; postoperative myocardial infarction was directed by the AHA's guidelines for determination of a myocardial infarction; and thromboembolic complications, falls, pressure sores, and iatrogenic pneumothorax were guided by the literature.

Second, we developed two computerized tools to improve incident data capture and analysis. The first tool, the Computerized Abstraction DEtection Tool (CADET), was developed for abstractors to use in adverse event identification and incorporated rule sets with criteria to define adverse event. The second tool, the Physician Event Review Kit (PERK), collated and distributed case summaries with an accompanied access database to collect and compare the adverse event ratings. Third, with the development of CADET, we utilized trained research nurses rather than physicians to conduct chart abstractions and prepare case summaries. Physicians were used to review incident case summaries created by the CADET from the nurses' chart abstraction findings. Case summaries were rated using PERK to determine the presence, preventability, and severity of suspected adverse events. We believed that this would be more cost effective than having physicians review charts, including many charts without actual events and excluded after nurse abstractions.

Table E2: Adverse events explicitly specified for identification during chart review

Adverse event type
Wound infection
Hospital-acquired urinary tract infection
Hospital-acquired pneumonia
Hospital-acquired bacteremia
Hospital-acquired sepsis
Operative nerve injury
Operative organ injury
Operative blood or lymph vessel injury or postoperative hemorrhage
Postoperative acute myocardial infarction
Postoperative stroke
Postoperative shock

Adverse event type
Postoperative respiratory distress or failure
Iatrogenic pneumothorax or hemothorax
Hospital-acquired pulmonary embolism
Hospital-acquired deep vein thrombosis
Fall
Pressure sore
Adverse drug event
Other adverse event not specified above

A significant difficulty with the assessment of the temporal relationship between working conditions and the occurrence of adverse events is accurate assignment of the time period during which the working conditions may have influenced adverse event occurrence. This time period is not necessarily the same as time period during which an adverse event is detected. In order to ensure the most likely time period relevant to when the adverse event occurrence was assigned by physicians, a set of rules was devised based upon the known temporal relationship between clinical identification of adverse events and the time of adverse event occurrence.

Reliability testing: We conducted reliability testing of our chart review method at three levels: the electronic administrative screening tool for chart selection, nurse chart abstraction, and physician event rating using case summaries. We abstracted a random sample of 1,990 screen negative charts. We found that 1,774 admissions had no adverse events (true negative) and that 216 admissions (11%) contained adverse events (false negative). Thus, the sensitivity and specificity of our screening tool were 0.44 and 0.75, respectively.

Using a random sample of screen positive charts, we tested the IRR among nurse abstractors and intra-rater reliability between nurses and physician researchers with experience in adverse event chart abstractions. We employed two sets of nurse chart abstractors, as one of the study hospitals was geographically distant from the other three hospitals. The level of intra-class correlation (ICC) between nurses was good to excellent (ICC = .61 and .96 at the central and distant study sites, respectively). Intra-rater reliability, conducted at the central study site, between nurses and physician was good (kappa = .69).

Last, we tested the reliability of physician rating for adverse event classification, severity, and preventability. Using random samples of event case summaries, our IRR for event classification (kappa = .69), event severity (kappa = .67), and preventability (kappa = .50) were comparable to prior adverse event studies using physician chart review and to prior adverse drug event studies using case summaries (as we did in this study).

Independent Variables

Individual patient risk: We controlled for certain patient characteristics that were known or suspected to be confounded with the likelihood of experiencing an AE. These included age, sex, and DRG weights. We used “adjacent DRG” categories to represent patient complexity that might predict individual risk of complications. DRGs that were separated by the presence or absence of comorbidities or complications were collapsed to avoid adjusting for the complication being measured. In other words, adjacent DRGs represent a risk adjustment method that does not adjust out patients’ risk

based on their actual complications. This same approach is used in the Patient Safety Indicator software distributed by AHRQ.[5]

Workload variables: We identified three measurable attributes of hospital activity or workload: *Census or occupancy rate, Throughput,* and *Intensity*.

Census/Occupancy Rates refer to the number of patients being cared for in the system at a given point in time. Because average daily census varies enormously from hospital to hospital, our primary variable was occupancy rate, defined as the number of patients per available (or operating) bed. High occupancy infers that the hospital is reaching its capacity. Most usual measures of census count the number of patients at midnight. However, this undercounts the workload pressure for high-occupancy hospitals that may discharge patients in the morning and admit a different patient at night. Because we were able to track patient locations by day, we accounted for this additional "occupancy" and referred to it as "true occupancy" for purposes of characterizing the four study sites. However, this resulted in occupancy rates greater than 100%, so, or analytic purposes, we calculated occupancy rate as the census divided by the 99th percentile of the census.

In addition, we measured when a hospital was on "diversion" as a proxy for when hospital capacity has been reached. Although diversion obviously is related to occupancy rates, the relationship is imperfect, because hospitals may divert patients when specific patient areas, such as the ICU, the catheterization lab, or the emergency department, are too crowded, even though hospital-wide occupancy may be less than 100%. Indeed, in preliminary analyses, diversion was not a significant predictor of AEs and so was dropped from the analysis.

Throughput is a second measure of workload and refers to turnover, or the rate at which patients move through the system during a specified time period. Two units with equal occupancy rates may have very different levels of admission/discharge turnover. In preliminary conversations with hospital nurses, we found that performing the services necessary for admitting or discharging patients requires far greater effort than caring for patients already in the unit, yet nurse staffing formulas do not take this activity into account. Thus, admission/discharge turnover, also referred to as "turnover ratios," or "activity ratios," may be equally important as a cause of AEs as census measures.

Intensity refers to the case complexity or severity of illness of the patient. Hospital workload increases with sicker patients. On each day, we calculated a severity-weighted census, defined as the sum of the DRG weights for all patients in the hospital on that day. For variable, we used the actual DRG, not the modified "adjacent" version.

We requested both hospital-level and unit-level workload information for each day of the study period. Unit-level data was only requested for medical and surgical units. The hospital-level request included the entire hospital (including all Ob/Gyn units, inpatient psychiatric units, etc.). Information for which hospital level is the sum total of the unit-level information, such as census, can be delivered as a single list of all the unit-level information. Because of difficulties in tracking specific units (units open and closing, changing names, and inconsistencies and disagreements with the discharge data), we collapsed all units into two "Superunits" – adult med-surg ICU and all other adult med-surg.

We also identified two "controllable" workload variables (weekday admission, emergent admission). These are controllable in the sense that hospitals have the ability to control the flow of scheduled admissions over selected days of the week.

Staffing data like workload data, were requested on a unit and a hospital level for each day of the study period. Although we explored numerous measures, in the end, we used the patient to nurse (PTN) ratio, because it has been shown in previous research to be related to patient safety.[7]

Patient Location: Patients are often moved between critical care and noncritical care settings during the course of their hospitalization, according to changing care needs. However, there is an element of uncertainty concerning the patient's ward location at the time of experiencing an adverse event. In view of this data limitation, we chose to concentrate our analyses upon the total medical and surgical ward location universe, including both ICU and non ICU locations, rather than attempt to model the relationship between ward-level workload conditions and adverse event frequency. Subanalyses were performed by grouping patients into total ICU and non-ICU locations; however, in view of the risk of patients being transferred to an ICU as a consequence of an adverse event, these analyses are not considered definitive evaluations of the relationship between ICU and non-ICU location workload conditions and the frequency of adverse events.

Analysis

The main focus of our analysis was the relationship between workload and staffing variables and adverse events. This analysis was conducted using Stata (Version 8, http://www.stata.com/). Analysis was conducted at all times under expert statistical supervision.

As noted previously, the study sample for the workload analysis included both screen-positive and screen-negative cases. This was a crucial step because of the possibility that cases were more likely to be screen positive on high-workload days. However, only screen-positive charts were reviewed for the presence of AEs (in supplemental work, a small subset of screen-negative cases were reviewed, but the number was small relative to the number of all screen-negative cases and so would not have affected our results in a meaningful way). For purposes of analysis, we assumed that screen-negative cases did not contain AEs. This was a conservative assumption, because, in subsequent work, about 10% of screen-negative cases were found to have AEs.

Our analytic strategy was based on the literature on patient safety and organizational behavior. The goal was to determine if adverse events were more likely to occur on peak workload days over the course of 1 year. We first characterized each day of the year according to its workload level, including occupancy rate, throughput, case complexity, and staffing; divided each measure into quartiles (representing 91 days); and then examined unadjusted daily rates of AEs per hospitalized patient. It is important to state what the study was not intended to do. This was not a test of whether AEs are more common at busier hospitals, because we only had four hospitals in the study. Rather, the focus was on day-to-day variation within a given hospital. Similarly, although we examined day-to-day variation in staffing, this was not a test of the association of staffing levels, *per se*, also because of the limited number of facilities.

Our main analysis employed the patient day as the unit of analysis. Our data showed that patients may have more than one AE per day, so we used Poisson regressions, which are particularly suited to counted data. Also, because the likelihood of a patient having an AE on any particular day is probably correlated to other days that they are in the hospital, we controlled for clustering by patient. We used the models to predict the likelihood of an adverse event occurring for a particular patient on a particular day. We estimated three types of models. Model 1 controlled for individual patient risk, including age, sex, and "adjacent" DRG categories. We included whether the patient was being treated in the ICU on the day of the AE, because the rate of AEs is high in ICUs.[8] A separate regression was estimated for each workload predictor, entered as a separate independent variable. Thus, five separate regressions were estimated. In model 2, we controlled in addition for certain confounding characteristics of the admission that are to some degree under the scheduling control of the hospital. These included the day of the week and whether the admission was emergent (i.e., not elective). Here again, we estimated five regressions, one for each workload variable. In model 3, we estimated a single regression, including all of the variables entered at once. These models were run for all four hospitals combined and for each hospital individually. When we ran the combined models, we constructed standardized workload variables to account for the differences in activity among the facilities.

Results (Principal Findings, Outcomes, Discussion, Conclusions, Significance, Implications).

The table shows census, "true" occupancy rates by percentile, super-unit, and day of the week. The true occupancy rates at HOSPITAL A are 97% at least half of the year, and the mean rate is 95%, which is considerably higher than the other hospitals.

Table E3 - Census and "True" Occupancy Rates by Percentiles and Day of Week (DOW), at Med-Surg Non-ICU and ICU Level

CENSUS	Non-ICU					ICU				
	A	**B**	**C**	**D**		**A**	**B**	**C**	**D**	
25th percentile	416	188	69	120		72	42	5	11	
50th percentile	451	205	78	130		76	47	7	13	
75th percentile	480	219	88	139		79	52	8	14	
99th percentile	513	245	121	157		87	60	12	17	
Mean daily	445	202	79	129		75	47	7	13	
"True" Occup Rates	**All Med Surg**					**ICU**				
	A	**B**	**C**	**D**		**A**	**B**	**C**	**D**	
25th percentile	90%	81%	55%	75%		80%	62%	42%	69%	
50th percentile	97%	89%	62%	81%		84%	69%	58%	81%	
75th percentile	102%	95%	70%	86%		88%	76%	67%	88%	
99th percentile	109%	106%	94%	97%		97%	88%	100%	106%	
Mean daily	95%	88%	63%	80%		83%	69%	57%	79%	

Table E4 contains the overall results of our chart review process. From among the 6,841 cases reviewed, 1,530 AEs were found by nurses and confirmed by the physician over-readers. Thus, the analytic study sample of 24,676 contained 1,530 cases with AEs and 23,146 cases without AEs (17,835 screen-negative cases plus 5,311 screen-positive cases found not to have AEs upon chart review).

Note that the screen-negative cases were assumed for purposes of analysis to not contain AEs. A small subsample of screen-negative reviews were performed to establish the sensitivity and specificity of our screens. They were not used in the analyses for the main analysis.

Table E4 - Study Sample, Numbers Screened, AEs Found, by Hospital

	A	B	C	D	All
Total eligible med-surg admissions	28,248	14,693	6,456	8,746	58,143
Sample of cases	8,499	7,414	3,866	4,897	24,676
screen positive cases (all were reviewed)	2,504	2,137	960	1,240	6,841
confirmed AEs from screen positive	721	482	120	207	1530
AE rates					
% of screen pos	28.8%	22.6%	12.5%	16.7%	22.4%
% of screen neg	10.2%	13.0%	9.0%	-	10.8%
% of entire sample	8.5%	6.5%	3.1%	4.2%	6.2%
# AE per patient-day	1.21%	1.08%	0.50%	0.68%	0.98%
#Patient-days for reviewed screen-positive patients	28,105	24,056	6,286	17,316	75,763
#Patient-days for reviewed screen-negative patients	26,117	18,298	12,207	9,880	66,502
Total # patient-days	54,222	42,354	18,493	27,196	142,265

Note: Only results from the screen-positive reviews were used as the basis for the workload study. Results from the screen-negative reviews were used to assess sensitivity and specificity for the supplemental studies.

Multivariate models

Due to space considerations, our unadjusted analyses are not presented here. Our original hypotheses assumed that we would find similar results across hospital. Even though the average workload differed, we thought that peak workload effects would occur regardless of facility. However, in unadjusted analyses, results differed greatly by hospital; furthermore, when we combined all the hospitals into a single model (standardizing the workload variables by subtracting the means and dividing by the sd), no workload effects were significant. Therefore, we present the results by individual hospital.

The Table contains the results of our multivariate models. The most apparent observation is that few of the results are significant except for Hospital A, the large teaching hospital with very high occupancy rates. For Hospital B, only the number of admissions was significant in model 1, and only discharges were significant in model 2, although it appears as if higher numbers of discharges were related to *fewer* AEs, which was contrary to our hypothesis. For Hospital C, there were no significant results for any of the models. For Hospital D, only the number of admissions was significant in model 1. Several of the coefficients for Hospitals B, C, and D were negative (but not significant), suggesting an inverse relationship between workload and AEs. Thus, it is unlikely that increased power would have supported our original hypotheses for these hospitals.

Hospital A, on the other hand, had positive and strongly significant results supporting the idea that increased workload was associated with an increased risk of AEs. All p values were .015 or less for model 1. Model 2 includes the individual risk variables (age, sex, DRG, and ICU) as well as what we call "controllable" workload variables (weekday admission, emergent admission). Each workload variable was entered in separate regressions for each hospital. For Hospital A, admissions and patients per nurse were significant at $p<.05$, and occupancy rate, discharges, and DRG-weighted census were significant at $p<.10$. Model 3 is the full model, containing all variables from models 1 and 2 entered into a single regression equation. These results should be interpreted in light of the high collinearity between all of the workload variables. Only patients/nurse was significant in this model, at $p<.10$. Occupancy rate had a negative but nonsignificant coefficient.

Table E6 - Results of Multivariate Models, Patient-Day Analysis

		Model 1			Model 2			Model 3		
HOSP		coeff	RR	P	coeff	RR	P	coeff	RR	P
	Occ rate	2.50	12.177	<.001	1.4325	4.19	0.08	-2.930	0.053	.140
A	Admissions	.0083	1.008	<.001	0.0054	1.01	0.0317	.0050	1.005	.135
	Discharges	.0076	1.008	<.001	0.0055	1.01	0.0638	.0035	1.004	.395
	Pnts/nurse	2.418	11.222	.015	2.4498	11.59	0.0173	2.798	16.411	.067
	DRG-weighted census	.0011	1.001	.001	0.0006	1.0006	0.0959	.0006	1.0006	.237
	Occ rate	.639	1.895	.253	-.842	0.431	.223	.392	1.48	.753
B	Admissions	.0107	1.011	<.001	.0004	1.0004	.933	.008	1.008	.204
	Discharges	-.0011	0.999	.804	-.0138	0.99	0.0112	-.016	0.984	.021
	Pnts/nurse*				---	---	---			
	DRG-weighted census	.0003	1.0003	.690	-.0011	0.999	.144	-.001	0.999	.451

			Model 1			Model 2			Model 3	
HOSP		coeff	RR	P	coeff	RR	P	coeff	RR	P
	Occ rate	-.2158	0.806	.818	-.6803	0.5065	.515	-1.400	0.49	.508
C	Admissions	.0228	1.023	.132	.0143	1.014	.460	.0325	1.033	.185
	Discharges	-.0036	0.996	.833	-.0138	0.986	.484	-.0055	0.995	.827
	Pnts/nurse*				---	---	---			
	DRG-weighted census	-.0012	0.999	.790	-.0034	0.999	.486	-.0014	0.999	.846
	Occ rate	2.447	11.553	.023	1.703	5.489	.218	3.544	34.617	.146
D	Admissions	.0154	1.015	.165	.0003	1.0003	.983	-.0098	0.990	.542
	Discharges	.0163	1.016	.110	.0021	1.002	.878	-.0064	0.994	.681
	Pnts/nurse	1.161	3.194	.260	.5352	1.708	.611	-.9984	0.368	.526
	DRG-weighted census	.004	1.004	.150	.0014	1.0014	.644	-.0009	0.999	.785
	Occ rate	2.106	8.214	<.001	1.2664	3.548	<.001	.171	1.186	.709
All	Admissions	.0073	1.007	<.001	.0050	1.005	<.001	.005	1.005	.025
	Discharges	.0066	1.0066	<.001	.0047	1.0047	<.001	-.0008	0.999	.789
	Pnts/nurse*	-.6041	0.547	<.001	-.5000	0.607	.001	--	--	--
	DRG-weighted census	.0003	1.0003	<.001	.0003	1.0003	<.001	.00001	1.00001	.927

NOTES:
* -- Patients per nurse could not be calculated for hospital B or C. Thus, models 1 and 2 for "all" hospitals only contains results for hospitals A and D combined. Furthermore, the results for model 3 for all hospitals contains all variables *except* patients/nurse. In model 1 – controlling for age, sex, adjacent DRGs, ICU – each workload variable is entered as a single independent variable in a separate regression. This will be five separate equations.
In model 2 – which controls for age, sex, adjacent DRGs, ICU, emergent admission, DOW – each workload variable is entered as a single independent variable in a separate regression. This will be five separate equations.
In model 3 – the "all variables model" – predictors included age, sex, adjacent DRGs, ICU, emergent admission, DOW, occupancy rate, admissions, discharges, num_weight, and patients per nurse. This is just one equation – one model.

Discussion

In an effort to compete in an increasingly cost conscious environment, hospitals pursued a number of strategies to limit costs and increase efficiency. Though some hospitals have closed, others now operate at or near maximum capacity, increasing patient-to-nurse ratios. Little is known, however, about whether these strategies affect patient safety. In the current study, we reviewed nearly 10,000 medical charts at four hospitals over the course of a single year.

We examined the association between the likelihood of adverse events and peak workloads. At three of the four hospitals, including one teaching hospital and two community hospitals, we could find no peak workload effect; at the fourth hospital (Hospital A), a major teaching hospital with very high ambient occupancy rates, the daily variation in number of admissions and patient-to-nurse ratios was strongly correlated with the occurrence of adverse events.

The explanation for this pattern of events is perhaps related to the concept of slack, as described by Perrow in his classic text, Normal Accident, in which he makes the point that tight coupling and high complexity (versus loose coupling and low complexity) are more accident prone.[9] Rudolph and Repenning[10] refer to this as slack and note that existing procedures work less well when there is no slack in the system, because the system loses its resilience to additional interruptions. As slack declines, coupling usually increases.

Our results have implications for patient safety in hospitals. Hospitals that operate at near capacity on a daily basis should consider re-engineering the processes of care to respond better during periods of high stress. These hospitals may wish to consider more closely the recommendations of the IOM, which include avoiding long shifts, simplifying key processes, standardizing work processes, creating systems to intercept or reverse errors before reaching the patient, and including patients in the care process.[2] Hospital administrators may decide to allow nursing supervisors more leeway in setting staffing levels or to institute policies that accommodate a larger on-call pool in order to flex up to the required number of nurses. At other hospitals, our results might imply that there is still excess capacity, at least from the standpoint of patient safety, and that enough slack resources are built into their structure so that they can function safely during periods of peak volume. Under such a scenario, very high volumes may be considered safe.

An obvious outcome of this study is its implications for future research. A likely follow-up to this study would be to perform research on the root causes of the errors that occur on busy days, investigating, for example, the interaction between crowding, coordination, and staff rotation on patient care units and their effects on AEs. However, it does not make sense to pursue such research until one is reasonably certain that such an association between workload and errors exists in more hospitals. Likewise, other research efforts may confidently address processes that function during average activity levels.

Aim Two: THE DETECTION OF ADVERSE EVENTS USING COMPUTERIZED METHODS

The attention given to the frequency and preventability of medical errors naturally suggests that hospitals monitor the occurrence of AEs in their institutions. However, the costs of doing so are prohibitive if only chart review methods are employed. Development of new methods to identify and track adverse events may save money and eventually improve quality.

Methods (Study Design, Data Sources/Collection, Interventions, Measures, Limitations).

Comparison of manual and computerized adverse event detection tools

Analyses were conducted consistent with our second major study aim of evaluating the accuracy and cost effectiveness of computerized administrative data screening adverse event detection tools in comparison to manual chart review methods and to evaluate the accuracy and cost effectiveness of Natural Language Processing as an adverse event screening technique when used to screen electronic hospital discharge summaries.

Patient specific data (including all ICD-9 codes) for all patients admitted to Hospital B between 1/10/2000 and 1/12/2001 (n = 30,710) were collected and sent for analysis by screening methods, including the Bates' method. The ICD-9 codes that were indicative of possible adverse events were used to select patients who were more likely to have experienced an adverse event during their hospitalization. A subset of those patients' charts were selected and used for manual chart review to identify actual adverse events. The manual chart review at Hospital B was performed by registered nurses specifically trained in the chart review and the coding process.

We also searched our patient archival database and calculated the actual patient census and patient turnover for each room during the study period. These data were combined with nurse staffing data, pharmacist-verified computerized adverse drug event data, and infection control practitioner-verified computerized hospital-acquired infection data to identify staff-to-patient crowding that may put patients at risk for adverse events. Computer methods were also used to calculate the average nursing acuity for each room every day during the study period.

Adverse drug events and specific hospital-acquired infections (bloodstream, urinary tract, wounds, respiratory tract) identified by the manual chart review were compared to the pharmacist-verified computerized adverse drug events and infection control practitioner-verified computerized hospital-acquired infections at Hospital B during the study period. This information was used to compare the computerized surveillance methods with the manual chart review. Adverse drug events and hospital-acquired infections identified by chart review and not identified by computer surveillance were analyzed to determine why they were not identified by the computer surveillance. The criteria used to identify those adverse events for the chart review was examined by a physician at Hospital B to determine the potential for inclusion in the computer surveillance logic for the identification of adverse drug events and hospital-acquired infections.

In addition, the following clinical observations were identified and collected by the physician for each ADE: causative agents, manifestations, and actions performed in response to the ADE. Additionally, contextual information was collected about each ADE, in order to determine when the event occurred relative to the inpatient stay. Last, information about the data source for each significant clinical observation was collected.

Development of the Natural Language Processing electronic hospital discharge summary Adverse Event screening model

As a supplement to the grant, we developed a Natural Language Processing (NLP) system capable of recognizing references to two types of adverse events (ADEs, hospital-acquired infections) from medical discharge summaries. To this end, we 1) altered a set of existing NLP tools so that they could be applied to this study, 2) developed a knowledge base (by training with exiting documents) to support this process, and 3) provided an initial estimate of the accuracy of this tool. The NLP tool upon which this project was based is called the Medical Probabilistic Language Understanding System (MPLUS).

Beginning in the Spring of 2004, nurses reviewed 500 additional patient charts identified to be negative by the Bates' screening tool. The nurses used a revised CADET tool to document the identified AEs. The NLP tool used the information that was gleaned from phrases or simple sentences in the electronic discharge summaries from the nurse-identified AE patients.

In the NLP system that Hospital B has been developing, concepts are derived from words or other concepts computationally, through the use of Bayesian Networks (BN). The modelling effort is to define the structure and content of the BN for each simple concept and then to develop a higher-level BN for which the simple concepts can be fused, forming a higher-level concept. The system learns by incorporating examples prepared for it. These examples are used to indicate typical groups of words and phrase that can be understood as a target concept. The system builds a probabilistic model of these relationships, which allows it to recognize the trained phrases and other, similar phrases in the future.

We chose to focus on Adverse Drug Events (ADE) as the initial type of adverse events modeled. This approach allowed us to explore the general characteristics of the NLP model that we needed to develop. Following development of the ADE model, additional adverse events were added to the NLP model.

A cornerstone of the approach that we used to identify adverse events is the use of a discourse model that can combine simple concepts from different parts of a discharge summary to derive a new concept. For example, in the recognition of an adverse drug event, one of the concepts is the delivery of a medication to a patient, and the second concept is the occurrence of an untoward medical event. The discourse model brings these together, tests whether the drug and the medical event represent cause and effect, and estimates the overall strength of the evidence of cause and effect.

Following development, testing, and training for ADE detection, we then tested a range of prototypes for the detection of a broader range of medical events and proceeded to train this generalized Adverse Event detection model.

Although we envisage more refinement of the NLP model over time, we chose to test the sensitivity, specificity, and predictive value of the NLP AE detection model using the manual record review approach as the gold standard following completion of generalized adverse event detection training.

Results (Principal Findings, Outcomes, Discussion, Conclusions, Significance, Implications).

Comparison with the EMR

Infections

During the study period, 362 bloodstream infections (BSIs) were identified by the computerized surveillance system at Hospital B. Of those, 16 (16/362 = 4.4%) BSIs were identified by both the computerized surveillance system and manual chart review from the subset of study patients. All the BSIs that were identified by manual chart review were detected by the computerized surveillance system. There were 47 BSIs from the study patients that were identified by computerized surveillance and verified by infection control that were not identified by manual chart review.

During the study period, 778 urinary tract infections (UTIs) were detected by the computerized surveillance system. Of those, 33 (33/778 = 4.2%) UTIs were identified by both the computerized surveillance system and by manual chart review from the subset of study patients. One UTI was identified by manual chart review and not by the computerized surveillance system. That UTI was identified by nurse chart review from text included in a dictated physician report. NLP methods were not included in the computer surveillance methods. (See Nosocomial Infections Tables below for breakdown of key phrases by document type and infection site for infections not identified by computer surveillance.) There were 100 UTIs from the study patients that were identified by computerized surveillance and verified by infection control that were not identified by manual chart review.

During the study period, 239 lower respiratory tract infections (RTIs) were detected by the computerized surveillance system. Of these, 30 (30/239 = 12.5%) RTIs were detected by both the computerized surveillance system and by manual chart review from the subset of study patients. Eight (8/239 = 3.3%) RTIs were detected by manual chart review and not by the computerized surveillance. All eight of those infections were identified by the chart review through text contained in physician dictated reports. There were 30 RTIs identified by computerized surveillance and verified by infection control that were not identified by manual chart review.

During the study period, 392 wound infections were detected by the computerized surveillance system. Of these, 71 (71/392 = 18.1%) were detected by both the computerized surveillance system and by manual chart review from the subset of study patients. Twenty-four (24/392 = 6.1%) wound infections were detected only by manual chart review and not by computerized surveillance. All of those were identified by text from physician-dictated reports. However, seven of those infections had positive microbiology culture results that could have been used by the computer surveillance logic. The specimen codes used by those cultures were not codes used by the computer logic to identify wound infections. Those codes can be added to the computer logic for future surveillance.

There were 35 wound infections from the study patients that were identified by computerized surveillance and verified by infection control that were not identified by manual chart review.

Thus, seven wounds were missed by computerized surveillance when study patients had positive culture results. The terms used for specimen and body site screening by computer surveillance were not included in those culture results. All other infections missed by computerized surveillance were identified by manual chart review through reading dictated reports (H&Ps, ER, disharge summaries, etc.) The type of dictated report and phrases pertinent to identification of each missed infection have been stored into Access tables. This information will be also used to develop natural language algorithms to detect future infections.

Adverse Drug Events

During the 14-month study period, 494 ADEs were identified by computer surveillance. The computer identified 88 patients with ADEs from the chart review study population. Of those, 13 were also identified by chart review, and 100 patients were identified to have ADEs that were not identified by computer surveillance.

Those 100 patients were identified by the chart review to have experienced a total of 122 different ADEs. The physician at Hospital B reviewed each of the 122 ADEs, according to the formal review criteria, to identify how the chart review identified each of the ADEs (text reports, medication orders, laboratory tests). For some of the ADEs, the physician did not find any information that indicated the presence of an ADE. Thus, each ADE was classified as TRUE POSITIVE, FALSE POSITIVE, or UNCERTAIN. The number of events that were found to be TRUE-POSITIVE ADEs by the physician-review was 97 (97/122=79.5%). The number of events that were found to be FALSE-POSITIVE ADEs by physician review was 19 (19/122=15.6%). Six of the 122 (6/122=4.9%) were categorized as UNCERTAIN after physician validation.

Detailed information for the 97 ADEs with chart documentation collected by the Hospital B physician is presented in the following tables. Eighty-five different patients experienced one ADE, and 12 patients experienced two ADEs. The information from this analysis will be used to update the new NLP methods to improve the computer ADE surveillance at Hospital B.

BY MANIFESTATION: The 97 ADEs not identified by computer surveillance included 126 different textual representations of clinical manifestations that could be used for NLP surveillance.

Results – Use of the NLP

Upon physician review, 445 (16.9%) of the 2,630 patients were demonstrated to have AEs. In the subset identified using the screening criteria, 389 of 2,137 (18.2%) were AE positive and, of the 493 screen-negative patients, 56 (11.4%) were AE positive. Among these were 134 patients with an ADE; 97 of 2,137 screen-positive patients had an ADE recorded, and 37 of 493 screen-negative patients were noted to have a corresponding adverse drug event.

Of these 2,630 patients, 2,108 had retrievable discharge summaries; 1,790 discharge summaries came from the Bates-positive patients, and 318 discharge summaries came from the Bates-negative subset.

The modified version of MPLUS was run against these documents. It categorized 259 discharge summaries as consistent with the presence of an ADE and 1,849 as not consistent with adverse drug effects. The sensitivity of the NLP system for ADEs was 0.246 (30 of 122 ADE-positive patients). The system's specificity was 0.897 (1,781 of 1,986 ADE-negative patients).

These statistics, however, are based on the mixture of screen-positive and screen-negative patients analyzed. To provide estimates applicable to the entire screened population, we adjusted the numbers to reflect the overall ratio of screen-positive and screen-negative patients seen. We therefore used the probabilities derived from the study to estimate rates that would have been seen in the initial cohort of 14,693 patients, of whom 7,414 were screened positive.

Discussion

A set of assumptions underlie this work. These are:

1) Adverse drug events can be expected to be represented by a series of concepts that appear separately in the medical narrative. These concepts are:

a. The administration of a medication.

b. The occurrence of an adverse medical event consistent with this medication.

c. Optionally, a clear statement about a response to the adverse drug event that describes the elimination of the medication, the delivery of antidotes, or both.

2) Neither the appearance of the medication order nor that of the adverse event caused is inadequate demonstration of an ADE.

The resulting NLP application therefore consisted of two embedded sub-applications, the first for identifying medications administered during the course of care, and the second designed to recognize a limited set of medical events consistent with an adverse response to these medications. In an attempt to increase the accuracy of the NLP system, the group of medical events modeled included some adverse events for which the description resemble that of medication-related illnesses but which generally occur independent of medications. Our experience suggests that including these competing concepts adds to the ability of the application to discriminate the target events.

A challenge in analyzing the success of this natural language processing application is that, though the physician review gives us a gold standard for the existence of ADEs in each patient, we do not know whether the ADEs noted by the reviewers are documented in the discharge summaries. In cases when reviewers confirmed an ADE based on a different document or inferred its existence by inspecting the medication record and lists of patient laboratory values or findings, the presence of an ADE may be recorded by the reviewer without it being documented in the discharge summary.

In general, the accuracy of the NLP system was disappointing. The overall sensitivity of the system was 24.6%, and the overall specificity was 89.47%. Interestingly, ADEs appear also to be a difficult area for the administrative screening criteria used. In this study, ADEs were more common (7.5%) in the screen-negative

group of patients than they were in the screen-positive group (4.5%). It may be that a combination of the two approaches (screening based on administrative data and natural language analysis of discharge reports), possibly augmented by other data from the electronic medical record, will prove more accurate than any single source of data.

References:

1. Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*. 2001, Washington, DC: National Academy Press. 364.
2. Institute of Medicine, *To err is human: Building a safer health system*. 2000, Washington, D.C.: National Academy Press.
3. Bates, D.W., et al., *Evaluation of screening criteria for adverse events in medical patients.* Medical Care, 1995. **33**(5): p. 452-62.
4. Iezzoni, L.I., et al., *Identifying complications of care using administrative data.* Medical Care, 1994. **32**(7): p. 700-15.
5. Miller, M.R., et al., *Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns.* Health Serv Res, 2001. **36**(6(Part II)): p. 110-132.
6. U.S. Centers for Disease Control and Prevention, *National Nosocomial Infections Surveillance System*. 2005.
7. Rothberg, M.B., et al., *Improving nurse-to-patient staffing ratios as a cost-effective safety intervention.* Med Care, 2005. **43**(8): p. 785-91.
8. Rothschild, J.M., et al., *The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care.* Crit Care Med, 2005. **33**(8): p. 1694-700.
9. Perrow, C., *Normal Accidents: Living With High-Risk Technologies*. 1984, New York: Basic Books.
10. Rudolph, J.W. and N.P. Repenning, *Disaster Dynamics: Understanding the Role of Quantity in Organizational Collapse.* Admin Sci Q, 2002. **47**: p. 1–30.

List of Publications and Products

Dr. Weissman and the project staff made the following presentations:

Published Abstracts And Presentations At Scientific Meetings

Bendavid E, Kaganova E, Rothschild J, Cook EF, Bates D, Weissman JS. Measures of Workload and Inpatient Complications Using AHRQ's Patient Safety Indicators in Three Massachusetts Hospitals. Agency for Healthcare Research and Quality 2nd Annual Patient Safety Conference, Arlington, VA, March 5, 2003.

Weissman JS, Bendavid E, Rothschild J, Cook EF, Bates D. Demonstration of the Computerized Adverse Event Detection Tool (CADET). Agency for Healthcare Research and Quality 2nd Annual Patient Safety Conference, Arlington, VA, March 5, 2003.

Weissman JS, Rothschild J, Cook F, Bendavid E, Bender M, Sprivulis P, David-Kasdan J, Kaganova J, Selbovitz L, Evans E, Haug P, Lloyd J, Bates D. The Relation of Crowded Working Conditions to Patient Safety in Hospitals. Annual Meeting of AcademyHealth. Boston, MA, 2005.

Invited Presentations

Weissman, Joel S. Weissman, Ph.D. "The Relation of Crowded Working Conditions to Patient Safety in Hospitals." 1st Annual Betsy Lehman Center for Patient Safety & Medical Error Reduction Symposium. Waltham, MA. December 2, 2004.

Weissman, Joel S. Weissman, Ph.D. "The effects of peak workload on hospital complication rates: A study of four hospitals." Boston University School of Management, Health Services Seminar. October 20, 2005.

Peer reviewed manuscripts

In addition to writing a final report for the agency, we are currently writing original articles for peer-reviewed journals. One overall piece will summarize the nature of the association between adverse events and crowding, if it exists. The executive summary contains a draft of the article. A second major article will detail the feasibility, accuracy, and cost of performing mechanized record reviews to monitor adverse events.