

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2011.01340.x
RESEARCH ARTICLE

The Sensitivity of Adverse Event Cost Estimates to Diagnostic Coding Error

Gavin Wardle, Walter P. Wodchis, Audrey Laporte, Geoffrey M. Anderson, and G. Ross Baker

Objective. To examine the impact of diagnostic coding error on estimates of hospital costs attributable to adverse events.

Data Sources. Original and reabstracted medical records of 9,670 complex medical and surgical admissions at 11 hospital corporations in Ontario from 2002 to 2004. Patient specific costs, not including physician payments, were retrieved from the Ontario Case Costing Initiative database.

Study Design. Adverse events were identified among the original and reabstracted records using ICD10-CA (Canadian adaptation of ICD10) codes flagged as postadmission complications. Propensity score matching and multivariate regression analysis were used to estimate the cost of the adverse events and to determine the sensitivity of cost estimates to diagnostic coding error.

Principal Findings. Estimates of the cost of the adverse events ranged from \$16,008 (metabolic derangement) to \$30,176 (upper gastrointestinal bleeding). Coding errors caused the total cost attributable to the adverse events to be underestimated by 16 percent. The impact of coding error on adverse event cost estimates was highly variable at the organizational level.

Conclusions. Estimates of adverse event costs are highly sensitive to coding error. Adverse event costs may be significantly underestimated if the likelihood of error is ignored.

Key Words. Medical errors, patient safety, hospital costs, propensity matching

A growing branch of research in the patient safety arena is the estimation of the economic cost of adverse events. Prior examinations have consistently reported alarmingly high estimates of the cost of adverse events (Kohn et al. 2000; Zhan and Miller 2003a; Zhan et al. 2006; Mello et al. 2007; Encinosa and Hellinger 2008). Not surprisingly, these estimates have caught the attention of policy makers who are increasingly incorporating adverse event cost estimates into hospital reimbursement schemes. The Centers for Medicare

and Medicaid Services (CMS) decision to withhold additional hospital payments for certain “conditions that could reasonably have been prevented” and “serious preventable events” has probably been the most debated policy response to date (Rosenthal 2007; Wachter, Foster, and Dudley 2008).

While chart review has been the dominant method for identifying adverse events in hospitalized patients (Thomas, Lipsitz, and Studdert 2002; Zhan and Miller 2003b; Michel, Quenon, and de Sarasqueta 2004), patient safety researchers are increasingly turning to administrative data (e.g., Rosen et al. 2006; Encinosa and Hellinger 2008; Houchens, Elixhauser, and Romano 2008; Raleigh et al. 2008; Rivard et al. 2008; Friedman et al. 2009). While the advantages of administrative data over chart reviews in terms of cost and coverage are considerable, previous studies have found that adverse events are incorrectly coded in administrative data compared to patient charts or other reference standard datasets (Best et al. 2002; Romano et al. 2002; Quan, Parsons, and Ghali 2004; Leibson et al. 2008). In Ontario, the setting for this study, previous analysis has demonstrated that diagnosis coding error in administrative data can have important implications for cost and case mix analysis (Preyra 2004; Sutherland and Botz 2006). While some researchers have acknowledged the likelihood of coding error and hypothesized that it would lead to an underestimation of the cost of adverse events (Zhan and Miller 2003a), there is, as yet, no published study that has estimated the impact of coding error on estimates of adverse event costs.

This article aims to address this gap in the literature by measuring the effect of coding error on estimates of the costs attributable to Needleman et al.’s (2002) nursing sensitive adverse events. These events include central nervous system complications, deep venous thrombosis, hospital-acquired pneumonia, hospital-acquired sepsis, metabolic derangement, pressure ulcers, pulmonary failure, shock or cardiac arrest, upper gastrointestinal bleeding, urinary tract infections, and wound infections. The analysis relied on a large reabstraction study of medical records from 11 Ontario hospital corporations. Coding discrepancies between the original and reabstracted records were

Address correspondence to Gavin Wardle, Ph.D., Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, 95 Bertmount Ave., Toronto, ON M4M 2X8; e-mail: gavin@preyrasolutions.com; gavin.wardle1@gmail.com. Walter P. Wodchis, Ph.D., Audrey Laporte, Ph.D., Geoffrey M. Anderson, M.M.D., Ph.D., and G. Ross Baker, Ph.D., are with the Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON. Walter P. Wodchis, Ph.D., is with the Toronto Rehabilitation Institute, Toronto, ON. Walter P. Wodchis, Ph.D., Audrey Laporte, Ph.D., and Geoffrey M. Anderson, M.M.D., Ph.D., are with the Institute for Clinical Evaluative Sciences, Toronto, ON.

examined and used to design a reference standard dataset that served as the basis for assessing the sensitivity of cost estimates to coding error.

DATA AND METHODS

The Discharge Abstract Database (DAD) is Ontario's acute inpatient administrative database and is maintained by the Canadian Institute for Health Information (CIHI). For each discharge, hospital-employed coders prepare a DAD abstract that includes one most responsible diagnosis (MRD_x) and up to 24 additional significant comorbidities. The coders also indicate whether each diagnosis was present on admission (POA) or manifested during the hospitalization. All diagnosis codes are abstracted using a Canadian adaptation of ICD10 (ICD10-CA).

Needleman et al.'s (2002) adverse events were identified based on the appropriate ICD10-CA codes that were flagged as postadmission complications. We selected Needleman et al.'s (2002) adverse events because they have been the subject of much research and the majority of the adverse events pertain to both medical and surgical patients. Other sets of adverse events, including AHRQ's Patient Safety Indicators, focus exclusively on surgical patients.

Canadian Institute for Health Information's coding standards state that diagnoses are to be abstracted only if the patient chart includes physician documentation that the condition satisfied at least one of the following criteria for significance: (1) the condition significantly affected the treatment received; (2) the condition required treatment beyond maintenance of the preexisting condition; or (3) the condition increased the length of stay by at least 24 hours (CIHI 2006). Although the standards aim to be specific, they may be open to multiple interpretations. For example, clinical evidence that attributes a 24-hour increase in a patient's hospital stay to a comorbid diagnosis may be ambiguous. Coders may therefore often be required to make a subjective assessment of the effect of diagnosis on a patient's hospital course. In the United States, CMS's guidelines for coding secondary diagnoses include similar criteria: "For reporting purposes the definition for 'other diagnoses' is interpreted as additional conditions that affect patient care in terms of requiring: clinical evaluation; or therapeutic treatment; or diagnostic procedures; or extended length of hospital stay; or increased nursing care and/or monitoring" (AMA 2008). Given the possibility for subjective coding decisions, it is relevant to consider whether applications of the data (e.g., for payment or performance measurement) might influence hospital coding practice.

Ontario's hospitals are nonprofit private corporations that receive funding from the provincial government equal to approximately 85 percent of their total operating expenses. Government funding is allocated using global budgets which are routinely adjusted for inflation, program changes, and population growth. At the time of the study, additional adjustments to the base budgets were made using a payment model based on hospitals' relative cost per adjusted discharge. Hospitals received shares of incremental funding in proportion to the difference between their actual and expected cost per adjusted discharge. Expected hospital costs were adjusted for teaching intensity, size, geography, and case mix. While hospitals had little ability to affect the other adjustment factors, they could influence their measured case mix by coding more aggressively. Under this model, hospitals that coded more adverse events would, *ceteris paribus*, increase their share of available funding. This aspect of Ontario's payment model was similar to that used by CMS for the PPS prior to its "never events" policy.

The Reabstraction Study

Looking to assess the adequacy of coding for its case mix based payment model, the Ontario Ministry of Health and Long Term Care (Ministry) partnered with CIHI to conduct a reabstraction study of 2002/03 and 2003/04 records from Ontario's 11 case costing hospital corporations. The corporations collectively operated 16 individual hospital sites, each of which collects patient-specific cost data using a standardized methodology. The cost data are submitted to the Ministry's Ontario Case Costing Initiative (OCCI) database. OCCI data have been the subject of numerous quality reviews and are used by the Ministry and CIHI to develop case mix systems. The costs captured in the OCCI include total direct (e.g., nursing, laboratory, pharmacy, imaging) and indirect costs. Indirect costs are those associated with administrative and support departments and each patient's share of these costs is determined using a standardized methodology (Ontario Ministry of Health and Long Term Care 2010). OCCI costs do not include physician payments. The 11 OCCI hospital corporations included in this study accounted for approximately 23 percent of Ontario's total discharges during the study years. All costs presented are in Canadian dollars.

The reabstraction study focused on records with multiple comorbidities. Specially trained coders (reviewers) recoded 13,803 abstracts directly from the original patient charts while blind to the original abstracts. After

completing each abstract, the reviewers compared their coding with the original coding and characterized the nature of each observed discrepancy using one of the following descriptions.

- 1 *Chart documentation*: Assigned when the reviewer believed that the discrepancy was due to information on the chart being missed by the original coder, or that the specificity of the original code was not adequately supported by information in the chart.
- 2 *Significance*: Assigned when the original and review coders agreed on the presence of a condition but disagreed whether the code met CIHI's criteria for significance.
- 3 *Standards*: Assigned when the reviewer believed that the original coder had reported or omitted a diagnosis contrary to CIHI's coding standards, other than those standards pertaining to the criteria for significance.
- 4 *Optional*: Assigned when reviewer coded differently than the original coder, but after review, the reviewer believed that the original coding was acceptable.

The Hybrid Dataset

The original data are potentially limited for adverse event cost analysis because they include adverse events that, according to the reviewers, had been coded without adequate supporting chart documentation or in violation of CIHI's standards. The reabstraction data may also be suboptimal for the analysis because they exclude adverse events identified by both the original and reviewing coders but deemed only by the reviewers to have had an insignificant impact on the patient's course.

To redress the limitations of the original and reabstraction data, we established a hybrid dataset. We made the hybrid dataset by including the following: (1) all adverse events coded by both the hospital coders and the reviewers; (2) all adverse events coded only by the reviewers; and (3) all the originally coded adverse events that were omitted by the reviewers because of "significance" and "optional" discrepancies. The hybrid data therefore exclude adverse events deemed by the reviewers to have been originally coded without adequate supporting chart documentation or in violation of coding standards. While we propose the hybrid data as our reference standard for the cost analysis, they are still subject to the inherent limitations of administrative data. These limitations include the potential for poor concordance of

ICD codes with physicians' notes, misinterpretation of physician notes by coders, and transcribing errors by physicians and coders. The hybrid data therefore do not establish a gold standard, but they are appropriate for our examination because they minimize coding inconsistencies associated with subjective interpretations of the coding standards.

Case–Control Matched Samples

Similar to previous adverse event costing studies, we mitigated confounding in our analysis by preprocessing our data using propensity score matching prior to conducting our parametric analysis (Bates et al. 1997; Classen et al. 1997; Zhan and Miller 2003a; Encinosa and Hellinger 2008; Rivard et al. 2008). We estimated each patient's probability of experiencing any of the adverse events using logit regressions that controlled for the covariates presented in Table 3. The covariates included the following patient characteristics deemed likely to have confounding effects on cost and likelihood of experiencing an adverse event: age, gender, urgent admission, and an indicator variable for each of three types of care: medical, surgical, and major surgical. As was done in previous adverse event costing studies, we included indicator variables for the Charlson chronic conditions (Charlson et al. 1987; Zhan and Miller 2003a; Encinosa and Hellinger 2008; Rivard et al. 2008). The regressions also controlled for the fixed effects of 19 major clinical categories (MCCs) which provide a general description of the body system or type of clinical condition associated with the primary cause of admission.¹ Because 2 years of data were pooled for the analysis, an indicator variable for the 2002/03 fiscal year was included to control for potential time trends across the years. The regressions also included indicator variables for the 11 hospital corporations to control for hospital level effects, including those potentially associated with teaching mission, size, and cost efficiency.

Matching of cases to controls was performed using a one-to-one nearest-neighbor matching algorithm. For each adverse event case, our matching algorithm selected a nonadverse event control within a 1 percent difference in risk of experiencing an adverse event. The 1 percent difference has been used in previous adverse event costing studies (Zhan and Miller 2003a) and was selected here after investigating the impact of higher and lower thresholds on the degree of balance in the resulting matched samples. Statistical analysis and matching were performed using *R* version 2.12 and its *MatchIt* package (Ho et al. 2007; Stuart et al. 2007).

In the three matched samples, we assessed balance using the percent improvement in difference in means for each covariate. This measure is defined as $((|a| - |b|)/|a|) \times 100$, where a is the difference in means between the adverse event and nonadverse event groups in the raw sample and b is the difference in the matched sample. Although they are widely used to assess balance, we did not use t -tests of differences in means because they can be misleading and should be avoided (Imai, Stuart, and King 2008). Our parametric analysis controlled for the potential of residual differences in the distribution of the covariates in the matched samples.

Estimating the Excess Cost of Adverse Events

We were interested in estimating the mean causal effect of the adverse events on patient cost averaged over all patients in the sample. This quantity is known in causal inference theory as the average treatment effect (ATE). In this section, we adopt the notation of Ho et al. (2007). Let $Y_i(1)$ be the cost that would be observed for patient i with an adverse event and characteristics X_b and $Y_i(0)$ be the cost without the adverse event. The ATE is then defined as:

$$\begin{aligned} \text{Excess Cost of each Adverse Event Type} &= \text{ATE} \\ &= \frac{1}{n} \sum_{i=1}^n E[Y_i(1) - Y_i(0)|X_i] \end{aligned}$$

where the summation over i refers to the matched sample.

After matching, we used a generalized linear model (GLM) with a Gamma distribution and logarithmic link function to regress patient cost on the variables in Table 3 and indicator variables for each adverse event type, while controlling for the fixed effects of 19 MCCs and the 11 hospital corporations. We selected this model over an OLS regression on the natural logarithm of cost and other families of the GLM class of models using Manning and Mullahy's (2001) algorithm.

We used the estimated coefficients from the regressions to predict $Y_i(1)$ and $Y_i(0)$ for each patient and adverse event type. We then calculated the ATE of each adverse event type by taking the sample mean of the difference in predicted costs. We made this analysis using each of the original, reabstraction, and hybrid datasets, and examined the sensitivity of cost estimates to coding error by comparing the estimated costs from the three datasets.

We performed additional analysis to examine the nature of the coding discrepancies in our data. For each adverse event type, we computed and

compared the ATE of four subsets of adverse events: (1) events coded by both the original and reviewing coders; (2) events coded by the original coders but deemed insignificant or optional by the reviewers; (3) events coded by the original coders but deemed improperly coded by the reviewers based on chart documentation or standards; and (4) events coded only by the reviewers. The ATEs were estimated from the original matched sample using the methods described above but with the 11 adverse event indicator variables in our GLM model replaced with 44 indicators for each adverse event–discrepancy type combination.

RESULTS

After applying Needleman et al.'s (2002) exclusion criteria, our raw sample consisted of 9,670 medical and surgical records in which no individual appeared twice. Records with multiple diagnosis codes indicating a single type of adverse event were deemed to have a single occurrence of that adverse event type. Records could have more than one type of adverse event, and this occurred in 8 percent (783) of records in the raw sample.

Coding Reliability

Table 1 shows that the original data included 3,620 adverse events, the reabstraction data included 2,586, and the hybrid data included 3,394. Table 1 also reports estimates of five agreement measures resulting from assessments of the original data against both the reabstraction and hybrid data. We present the results of both assessments because these estimates, in conjunction with the cost estimates presented subsequently, may help other researchers gauge how coding accuracy in their jurisdictions might affect estimates of adverse event costs.

When assessed against the reabstraction data, the median sensitivity of original adverse event coding was 0.67 and the median PPV was 0.52. Adverse events with the highest sensitivities included urinary tract infections (0.78) and shock or cardiac arrest (0.76). Adverse events with the lowest sensitivities were upper gastrointestinal bleeding (0.60) and pulmonary failure (0.62). Agreement measures are higher for all adverse events when the original data are compared against the hybrid data. The median sensitivity was 0.76 and the median PPV was 0.73. Metabolic derangement, central nervous system complications, and urinary tract infections have the largest differences

Table 1: Agreement Rates for Adverse Event Codes

Adverse Event	Number of Cases with Adverse Event				Reabstraction Data as Reference Standard				Hybrid Data as Reference Standard					
	Original	Reab	Hybrid	κ	Sensitivity	PPV	Specificity	NPV	κ	Sensitivity	PPV	Specificity	NPV	κ
	Central nervous system complications	131	77	127	64.9	38.2	99.2	99.7	47.6	78.7	76.3	99.7	99.7	99.7
Deep venous thrombosis	173	162	171	71.0	66.5	99.4	99.5	68.1	72.5	71.7	99.5	99.5	99.5	71.6
Hospital-acquired pneumonia	400	285	312	73.3	52.3	98.0	99.2	59.6	75.6	59.0	98.2	98.2	99.2	65.0
Hospital-acquired sepsis	192	178	188	63.5	58.9	99.2	99.3	60.3	65.4	64.1	99.3	99.3	99.3	64.0
Metabolic derangement	801	328	704	64.6	26.5	93.7	98.7	34.4	83.5	73.4	97.6	97.6	98.7	76.3
Pressure ulcers	163	120	164	72.5	53.4	99.2	99.7	60.9	79.9	80.4	99.7	99.7	99.7	79.8
Pulmonary failure	164	91	114	61.5	34.1	98.9	99.6	43.2	69.3	48.2	99.1	99.1	99.6	56.2
Shock or cardiac arrest	665	597	642	76.2	68.4	97.7	98.4	70.2	77.9	75.2	98.2	98.2	98.4	74.8
Upper gastrointestinal bleeding	123	96	118	60.4	47.2	99.3	99.6	52.4	67.8	65.0	99.5	99.5	99.6	66.0
Urinary tract infection	523	347	529	77.8	51.6	97.3	99.2	60.4	85.4	86.4	99.2	99.2	99.2	85.1
Wound infection	285	305	325	66.6	71.2	99.1	98.9	67.8	68.6	78.2	99.3	99.3	98.9	72.2
Totals/median	3,620	2,586	3,394	66.6	52.3	99.1	99.3	60.3	75.6	73.4	99.3	99.3	99.3	72.2

Note. κ, Kappa statistic; NPV, negative predictive value; PPV, positive predictive value; Reab, reabstraction dataset.

in PPVs between the reabstraction and hybrid estimates (0.48, 0.38, and 0.35). Because they are largely driven by “significance” discrepancies, these differences indicate that coders had the most difficulty in assessing the effect of these conditions on a patient’s hospital course. Not shown here, the median sensitivity of the Charlson comorbidities was 0.85 and the median PPV was 0.81.

Table 2 shows transition matrices for diagnosis codes captured in the original and reabstraction data. There are 3,949 original adverse event codes in this table compared to the 3,620 adverse events reported elsewhere in the analysis because some records had multiple codes indicating the presence of a single adverse event type. These tables suggest two promising findings. First,

Table 2: Transition Matrix for Original and Reabstracted Diagnosis Codes

<i>Reabstracted as:</i>						
<i>Originally Coded as:</i>	<i>MRDx (%)</i>	<i>Present on Admission (%)</i>	<i>Postadmission (%)</i>	<i>Secondary* (%)</i>	<i>Not Re-Coded (%)</i>	<i>Total Diagnosis Codes</i>
MRDx	82	7	1	2	8	9,670
Charlson comorbidities	6	48	1	6	39	7,069
Other present on admission	4	42	3	2	49	18,235
Adverse events	0	4	49	1	45	3,949 [†]
Other postadmission	0	4	47	1	47	10,192
<i>Originally Coded as:</i>						
<i>Reabstraction Coding</i>	<i>MRDx (%)</i>	<i>Present on Admission (%)</i>	<i>Postadmission (%)</i>	<i>Secondary* (%)</i>	<i>Not Originally Coded (%)</i>	<i>Total Diagnosis Codes</i>
MRDx	82	11	1	1	5	9,670
Charlson comorbidities	5	76	2	5	12	4,511
Other present on admission	5	71	5	3	17	10,872
Adverse events	1	7	71	2	20	2,761 [†]
Other postadmission	0	6	69	1	23	6,985

*Secondary diagnoses are those that were present but were deemed not to have impacted the patient’s course of care. Given that they are optional for hospitals to code, they were beyond the scope of the reabstraction study.

†Due to occurrences where a single record has multiple diagnosis codes indicating the same type of adverse event, the adverse event counts herein are higher than the counts presented in previous tables.

the MRDx appears reliably coded; coders agreed on the MRDx in 82 percent of records. Second, if they agreed on the presence and significance of the diagnosis, coders reliably established the timing of diagnosis onset. For example, only 1 percent of originally coded Charlson comorbidities was reclassified as having manifested during the hospitalization and only 4 percent of originally coded adverse events were reclassified as having been POA.

Less promising is that coders had great trouble agreeing on the presence and significance of adverse events and comorbidities. The reviewers agreed with the code selection and typing of only 49 percent of the 3,949 originally coded adverse events. Moreover, the reviewers re-coded only 48 percent of the original Charlson comorbidities. The proportion of original codes that were not re-coded by the reviewers, shown in the second last column of Table 2, demonstrates the conservatism of the reviewers relative to the original abstractors. For example, the reviewers did not re-code 45 percent of the original adverse events. Not shown in the table, the reviewers deemed that these adverse events did not meet the criteria for significance (22 percent), had inadequate supporting documentation in the chart (15 percent), did not meet coding standards (7 percent), or were optional/not wrong to code (1 percent). The adverse events associated with significance and optional disagreements are excluded from the reabstraction data but ought to be included in the cost analysis because there is agreement among the coders on the presence of the adverse event. In contrast, the adverse events associated with disagreements over documentation and standards should be excluded because the reviewers disagreed with the original coders over the presence of the adverse events.

Despite their relative conservatism, the reviewers did code conditions that had been overlooked by the original coders. Shown in the lower section of Table 2, 20 percent of the 2,761 adverse events coded by the reviewers had not been captured in the original abstracts. The reviewers deemed that these adverse events were originally omitted due to information on the chart being missed by the original coders (15 percent), incorrectly deemed insignificant by the original coders (2 percent), or omitted in contravention of the coding standards (2 percent). These adverse events are apparent false negatives and should be included in the cost analysis.

The Matched Samples

Table 3 shows that the adverse event cases had different characteristics than the nonadverse event cases in the raw sample. Using our algorithm, we matched 2,194 adverse event records from the original data to 2,194

Table 3: Patient Characteristics in the Raw Sample and Original Matched Sample

	Raw Sample		Matched Sample		Percent Improvement in Difference in Means
	Adverse Event	No Adverse Event	Adverse Event	No Adverse Event	
Estimated probability of an adverse event	.316	.224	.303	.303	99.8
Patient characteristics					
Age	65.4	63.3	65.1	65.4	87.7
Female	47.3%	45.8%	46.8%	47.1%	81.6
Urgent admission	73.2%	74.3%	72.8%	72.9%	92.1
Medical admission	29.4%	50.2%	30.9%	31.5%	97.2
Major surgical admission	36.6%	21.8%	35.6%	35.9%	98.2
Year 2002	50.9%	51.7%	50.8%	49.7%	-40.5
Chronic conditions					
Myocardial infarction	4.1%	3.1%	3.9%	4.1%	85.8
Congestive heart failure	9.4%	8.3%	9.1%	9.6%	50.1
Peripheral vascular disease	3.5%	1.6%	2.9%	2.7%	90.3
Cerebrovascular disease	2.9%	1.7%	2.6%	2.7%	87.9
Dementia	3.1%	2.2%	2.9%	3.4%	41.8
Chronic obstructive pulmonary disease	8.3%	5.3%	7.0%	6.6%	86.0
Rheumatoid arthritis	1.7%	1.0%	1.4%	1.4%	94.1
Peptic ulcer disease	1.5%	0.8%	1.1%	1.1%	93.5
Mild liver disease	1.0%	0.8%	1.0%	0.6%	-38.4
Diabetes without complications	14.1%	10.3%	13.4%	14.1%	80.7
Diabetes with complications	6.5%	5.5%	6.2%	5.9%	70.2
Hemiplegia	3.3%	2.1%	3.2%	3.3%	88.8
Moderate liver disease	6.1%	4.9%	5.7%	5.7%	96.0
Cancer	7.8%	5.1%	7.4%	6.7%	73.2
Severe liver disease	1.2%	0.8%	1.2%	0.7%	-17.0
Carcinoma	8.7%	7.2%	8.5%	9.1%	55.7

continued

Table 3. Continued

	Raw Sample		Matched Sample		Percent Improvement in Difference in Means
	Adverse Event	No Adverse Event	Adverse Event	No Adverse Event	
Hospitals					
Hospital A	0.7%	2.4%	0.8%	1.0%	83.5
Hospital B	6.2%	7.5%	6.4%	5.7%	44.5
Hospital C	4.6%	3.9%	4.5%	4.4%	87.8
Hospital D	12.7%	9.7%	12.4%	11.9%	86.4
Hospital E	10.4%	7.1%	9.9%	10.9%	68.2
Hospital F	10.7%	8.9%	10.7%	10.1%	64.3
Hospital G	7.1%	16.0%	7.5%	7.3%	98.5
Hospital H	11.4%	9.1%	11.4%	11.1%	88.4
Hospital I	11.0%	8.2%	10.7%	11.4%	76.8
Hospital J	12.4%	10.3%	12.6%	13.5%	59.0
Hospital K	12.8%	17.1%	13.2%	12.7%	89.3
Mean cost (SE)	\$44,532 (1,015)	\$13,765 (231)	\$43,085 (1,021)	\$17,781 (490)	
N	2,309	7,047	2,194	2,194	

nonadverse event records on the basis of similarity in propensity scores. Matches that met our threshold of a maximum of 1 percent difference in predicted risk could not be found for 115 (5 percent) adverse event records. The last column of Table 3 shows that our matching algorithm improved the extent of balance between the case and control groups for all covariates except *Year 2002*, *Mild Liver Disease*, and *Severe Liver Disease*. That the mean propensity score was equal in the case and control groups (0.303) after matching indicates the overall success of our matching exercise. Similar improvements in balance were achieved for the reabstraction and hybrid matched samples.

Excess Costs

Table 4 shows the excess unit cost of each adverse event estimated from the original, reabstraction, and hybrid datasets. The mean excess cost was \$17,218 in the original data, \$26,157 in the reabstraction data, and \$22,642 in the hybrid data. The mean cost of all cases in the raw sample was \$21,358, which reflects the focus of the reabstraction study on complex cases. The mean cost of all cases at the OCCI hospitals was approximately \$7,000 in 2003/04.

Excess unit costs estimates derived from the original data were, on average, 24 percent less than the estimates derived from the hybrid data. Upper gastrointestinal bleeding events resulted in the highest excess costs (\$30,176) while metabolic derangements results in the lowest excess costs (\$16,008). The difference between the original and hybrid estimates was largest for deep venous thrombosis (41 percent).

The last eight columns of Table 4 show the counts and ATEs of each type of coding discrepancy. Adverse events listed under the “significance” column are those that were originally coded and subsequently deemed by the reviewers to have occurred but not to have satisfied CIHI’s criteria for significance. It would therefore be intuitive to expect that the ATEs of these events would be positive but less than the ATEs of events coded by both coders, and this was the case for 8 of the 11 adverse event types. On average, the ATE of insignificant events was 31 percent less than the average ATE of events coded by both coders. The ATE of insignificant occurrences of sepsis was negative. Contrary to intuition, the ATEs of insignificant occurrences of thrombosis, shock or cardiac arrest, and upper gastrointestinal bleeding were higher than the ATEs of occurrences coded by both coders. Given that the reviewers did not find adequate information in the charts to support coding these events, it is curious that the ATE for all events associated with chart documentation and standard discrepancies, save pressure ulcers, were positive. However, the

Table 4: Estimated Excess Unit and Total Costs from the Original, Reabstraction (Reab), and Hybrid Datasets

Adverse Event	Adverse Event Coding Discrepancies According to Reviewers											
	Incremental Cost (\$) SE			Original Cost Compared to Hybrid Cost (%)	No Discrepancy		Significance		Chart Documentation and Standards		Coded Only by Reviewers	
	Original	Reab	Hybrid		Count	Cost (\$)	Count	Cost (\$)	Count	Cost (\$)	Count	Cost (\$)
	Incremental	Reab	Hybrid	Count	Cost (\$)	Count	Cost (\$)	Count	Cost (\$)	Count	Cost (\$)	
Central nervous system complications	12,427 227	25,663 730	17,428 439	-29	50	22,888	50	2,992	31	14,771	27	12,918
Deep venous thrombosis	13,949 250	24,648 687	23,778 590	-41	115	19,586	9	22,779	49	10,061	47	18,372
Hospital-acquired pneumonia	20,817 340	27,313 593	22,536 479	-8	209	23,942	27	9,459	164	18,354	76	32,959
Hospital-acquired sepsis	22,883 382	34,491 856	26,143 622	-12	113	29,373	10	-2,572	69	12,330	65	12,365
Metabolic derangement	11,954 214	20,315 507	16,008 353	-25	212	20,183	376	12,325	213	5,660	116	13,038
Pressure ulcers	21,253 370	34,593 902	28,676 703	-26	87	28,058	44	15,615	32	-301	33	31,210
Pulmonary failure	15,012 270	31,467 795	22,483 534	-33	56	18,820	23	4,858	85	9,301	35	18,711
Shock or cardiac arrest	17,629 301	21,082 525	21,438 477	-18	455	17,124	45	33,830	165	14,178	142	33,047

continued

Table 4. *Continued*

<i>Adverse Event</i>	<i>Adverse Event Coding Discrepancies According to Reviewers</i>											
	<i>Incremental Cost (\$) SE</i>			<i>Original Cost Compared to Hybrid Cost (%)</i>	<i>No Discrepancy</i>		<i>Significance</i>		<i>Chart Documentation and Standards</i>		<i>Coded Only by Reviewers</i>	
	<i>Original</i>	<i>Reab</i>	<i>Hybrid</i>		<i>Count</i>	<i>Incremental Cost (\$)</i>	<i>Count</i>	<i>Incremental Cost (\$)</i>	<i>Count</i>	<i>Incremental Cost (\$)</i>	<i>Count</i>	<i>Incremental Cost (\$)</i>
Upper gastrointestinal bleeding	23,521 415	26,568 666	30,176 661	-22	58	20,202	22	35,170	43	14,662	38	42,743
Urinary tract infection	17,910 303	21,801 495	20,286 421	-12	270	20,748	182	17,271	71	24,392	77	19,950
Wound infection	12,040 225	19,789 513	20,109 478	-40	203	15,034	20	12,068	62	2,287	102	11,826
Mean/total	17,218	26,157	22,642	-24	1,828	21,451	808	14,890	984	11,427	758	22,467

average ATE of events with these discrepancies was 48 percent less than the average ATE of events coded by both coders. The ATEs of all events coded only by the reviewers were positive and, on average, 5 percent higher than the average ATE of events coded by both coders.

Institutional Level Results

Table 5 shows results at the institutional level. Compared to the hybrid data, adverse events were, on average, 7 percent over-reported in the original data with a range across the hospitals of 26 percent over-reported (Hospital D) to 27 percent under-reported (Hospital A). Institutional estimates of the total cost attributable to the adverse events were derived by multiplying the number of adverse events at each hospital in each dataset by the corresponding unit cost estimate. Among the four largest institutions in the study (H, D, J, F), the extent to which the total costs attributable to adverse events varied between the original and hybrid estimates ranged from 3 percent underestimated (Hospital D) to 25 percent underestimated (Hospital F). Not shown in the table, the original total cost estimates were less than the hybrid estimates by 14 percent on average for the teaching hospitals and 20 percent on average for the large community hospitals. The last row of Table 5 shows that the estimate of the total cost attributable to adverse events in the original data was 16 percent less than the estimate derived from the hybrid data.

DISCUSSION

Estimates of the excess unit and total costs attributable to adverse events are highly sensitive to diagnosis coding error. Coding error in the original data caused the excess unit costs to be underestimated on average, relative to our reference standard estimate, by 24 percent, and the total cost attributable to the adverse events to be underestimated by 16 percent. This is an important result because it suggests that the economic impact of adverse events might be under-estimated in studies that ignore the likelihood of error. Given this finding, previous assessments of the business case for patient safety may have been biased against the cost effectiveness of patient safety improvements.

The observed extent of institutional-level variation in adverse event coding and costs estimates suggests that that Ontario's administrative data form an inconsistent basis for hospital performance measures related to adverse events. The variation also suggests that hospital payment schemes based on

Table 5: Adverse Events and Estimated Excess Costs by Hospital

Institution	Hospital Type*	Hospital Sites	Adverse Events			Original Count Relative to Hybrid		Total Cost (\$) (unit cost times prevalence)			Original Total Cost Estimate Relative to Hybrid Estimate (%)
			Original	Reab	Hybrid	Original Count (%)	Hybrid Count (%)	Original	Reab	Hybrid	
			Original	Reab	Hybrid	Original	Reab	Hybrid	Original	Reab	
Hospital A	Small community	1	19	24	26	-27	278,943	524,713	528,291	-47	
Hospital B	Large community	1	195	178	206	-5	3,436,995	4,394,555	4,542,973	-24	
Hospital C	Large community	1	166	99	162	2	2,651,166	2,426,799	3,378,113	-22	
Hospital D	Teaching	1	482	280	384	26	7,777,068	6,491,362	8,009,132	-3	
Hospital E	Teaching	1	396	274	359	10	6,450,421	6,768,885	7,551,200	-15	
Hospital F	Teaching	1	436	378	454	-4	7,113,488	9,106,827	9,470,632	-25	
Hospital G	Large community	4	199	151	177	12	3,401,612	3,538,433	3,745,885	-9	
Hospital H	Teaching	1	501	271	456	10	8,312,911	6,636,840	9,495,749	-12	
Hospital I	Large community	1	394	336	400	-2	6,324,621	7,840,794	8,314,118	-24	
Hospital J	Teaching	1	442	280	391	13	7,099,322	6,964,454	8,143,203	-13	
Hospital K	Large community	3	390	315	379	3	6,830,487	7,613,998	8,172,757	-16	
Total		16	3,620	2,586	3,394	7	59,677,035	62,307,659	71,352,051	-16	

*Small hospitals are those with <3,500 cost-adjusted discharges per year, referral populations of <20,000 people, and are generally the single provider in a community. Teaching hospitals are designated members of the Council of Academic Hospitals of Ontario, and large community hospitals are all other short-stay acute hospitals.
Reab, reabstraction dataset.

these inconsistent measures could be unjust and lead to the misdirection of efforts to improve quality and contain costs. These findings may have important implications for jurisdictions considering the implementation of hospital reimbursement systems that rely on administrative data to identify and estimate the cost of adverse events.

We believe that the finding that the ATEs of events associated with significance discrepancies were positive for all events types except sepsis, coupled with the implicit agreement among the coders on the occurrence of the events, supports their inclusion in our reference standard hybrid data. Moreover, despite finding that they had positive ATEs, we believe that the explicit disagreement among the coders on the occurrence of events associated with chart documentation and standards discrepancies supports their exclusion from our hybrid data. A reasonable alternative to our approach might be to use only the events coded by both the original and review coders as the reference standard. The results of this analysis are consistent with our findings; ATE estimates derived from the original data were, on average, 19 percent lower than estimates derived using only the events captured by both sets of coders.

Our findings suggest a need to critically review the reliability of coding standards pertaining to adverse events. A noteworthy difference between CIHI and CMS standards is that CIHI requires that a diagnosis extended the hospital stay by at least 24 hours, whereas CMS requires only the extension of the hospital stay. Given the potential difficulties in attributing a specific length of stay increase to a particular diagnosis code, it might be worthwhile to investigate whether CIHI's minimum threshold increases the requirement for subjective coding. Moreover, our findings related to the extent and ATEs of events coded only by the reviewers (i.e., those apparently missed by the original coders) also points to a need to examine more upstream processes associated with chart documentation. It is possible that many of the "missed" events were found by the reviewers due to information added to the charts after the preparation of the original abstracts.

There may be opportunities to improve on adverse event identifying algorithms in the absence of reforms to coding and data collection mechanisms. While procedure codes are not relevant for all adverse events, they have been used to augment the identification criteria for some adverse events (Wahl et al. 2010) and are generally coded more reliably than diagnoses (Juurink et al. 2006). Results of studies that have enhanced administrative data with objective laboratory data for risk adjustment have been promising and suggest that similar approaches may be relevant for adverse event identification

(Pine et al. 2007; Tabak et al. 2010). The potential of pharmacy data to identify clinical interventions associated with the management of adverse events should also be explored.

Adverse event coding in our sample of Ontario data appears to be at least as reliable as that in U.S. jurisdictions. Using the National Surgical Quality Improvement Program data to assess coding in the Department of Veteran Affairs' Patient Treatment File, Best et al. (2002) found that only 7 percent of adverse event codes had sensitivities above 0.50, and only 4 percent of codes had positive predictive values above 0.50. Romano et al. (2002) used chart reviews to assess the quality of coding of postoperative complications among diskectomy patients in California and found that only 4 of 31 complications had sensitivities above 0.60. Looking to validate the complications included in the Complications Screening Program (Iezzoni et al. 1994); McCarthy et al. (2000) found that postoperative acute myocardial infarctions were well reported, but that <60 percent of other complications had adequate clinical evidence in the patient charts to support the diagnosis.

The costs reported herein are higher than those reported in previous studies. This may be due to differences in the costs being analyzed: we analyzed costs as reported by the treating hospitals, whereas previous articles have analyzed transacted payments (Encinosa and Hellinger 2008), charges (Zhan and Miller 2003a), or estimated costs based on hospital level cost-to-charge ratios applied to patient-level charges (Bates et al. 1997). Given these differences, jurisdictions outside Ontario may find the extent of variation in cost estimates (i.e., coding-error induced variation) of more interest than the point estimates. The differences in costs may also be due to the sampled patients. Our study focused on complex patients, whereas other studies investigated the impact of adverse events on cost for patients across all severity levels. No study has yet investigated whether the causal effect of adverse events on cost is constant across patient severities, but this would be a useful contribution to the literature.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was approved by the University of Toronto Ethics Review Board, and it was supported by a grant from the Canadian Institute for Health Research (CIHR grant 84310).

Disclosures: None.

Disclaimers: None.

NOTE

1. Major clinical categories are referred to throughout this document and are registered trademarks of the CIHI.

REFERENCES

- American Medical Association. 2008. "ICD-9-CM Official Guidelines for Coding and Reporting." Available at http://www.ama-assn.org/resources/doc/cpt/icd9cm_coding_guidelines_08_09_full.pdf
- Bates, D., N. Spell, D. Cullen, E. Burdick, N. Laird, L. Petersen, S. Small, B. Sweitzer, L. L. Leape, F. Name, D. W. Bates, C. Spell, L. Burdick, S. Petersen, and L. Sweitzer. 1997. "The Costs of Adverse Drug Events in Hospitalized Patients. Adverse Drug Events Prevention Study Group." *Journal of the American Medical Association* 277: 307–11.
- Best, W., S. Khuri, M. Phelan, K. Hur, W. Henderson, J. Demakis, and J. Daley. 2002. "Identifying Patient Preoperative Risk Factors and Postoperative Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program." *Journal of the American College of Surgeons* 194: 257–66.
- Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40: 373–84.
- CIHI. 2006. *Canadian Coding Standards for ICD-10-CA and CCI for 2006*. Ottawa, Canada: The Canadian Institute for Health Information.
- Classen, D., S. Pestotnik, R. Evans, J. Lloyd, and J. Burke. 1997. "Adverse Drug Events in Hospitalized Patients. Excess Length of Stay, Extra Costs, and Attributable Mortality." *Journal of the American Medical Association* 277: 301–6.
- Encinosa, W., and F. Hellinger. 2008. "The Impact of Medical Errors on Ninety-Day Costs and Outcomes: An Examination of Surgical Patients." *Health Services Research* 43: 2067–85.
- Friedman, B., W. Encinosa, H. J. Jiang, and R. Mutter. 2009. "Do Patient Safety Events Increase Readmissions?" *Medical Care* 47: 583–90.
- Ho, D., K. Imai, G. King, and E. A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15: 199–236.
- Houchens, R., A. Elixhauser, and P. Romano. 2008. "How Often Are Potential Patient Safety Events Present on Admission?" *Joint Commission Journal on Quality & Patient Safety* 34: 154–63.
- Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, E. S. Fisher, C. Duncan, J. S. Hughes, and G. A. Coffman. 1994. "Identifying Complications of Care Using Administrative Data." *Medical Care* 32: 700–15.

- Imai, K., G. King, and E. A. Stuart. 2008. "Misunderstandings Between Experimentalists and Observationalists About Causal Inference." *Journal of the Royal Statistical Society Series A* 171 (part 2): 481–502.
- Juurlink, D., R. Croxford, A. Chong, P. Austin, J. Tu, and A. Laupacis. 2006. *Canadian Institute for Health Information Discharge Abstract Database: A Validation Study*. Toronto, ON: Institute for Clinical Evaluative Sciences.
- Kohn, L., J. Corrigan, M. Donaldson, and Committee on Quality of Health Care in America, Institute of Medicine. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press.
- Leibson, C., J. Needleman, P. Buerhaus, J. Heit, L. Melton, J. Naessens, K. Bailey, T. Petterson, J. Ransom, and M. Harris. 2008. "Identifying In-Hospital Venous Thromboembolism (VTE): A Comparison of Claims-Based Approaches with the Rochester Epidemiology Project VTE Cohort." *Medical Care* 46: 127–32.
- Manning, W. G., and J. Mullahy. 2001. "Estimating Log Models: To Transform or Not Transform." *Journal of Health Economics* 20: 461–94.
- McCarthy, E. P., L. I. Iezzoni, R. B. Davis, R. H. Palmer, M. Cahalane, M. B. Hamel, K. Mukamal, R. S. Phillips, and D. T. Davies Jr. 2000. "Does Clinical Evidence Support ICD-9-CM Diagnosis Coding of Complications?" *Medical Care* 38: 868–76.
- Mello, M. M., D. M. Studdert, E. J. Thomas, C. S. Yoon, and T. A. Brennan. 2007. "Who Pays for Medical Errors? An Analysis of Adverse Event Costs, the Medical Liability System, and Incentives for Patient Safety Improvement." *Journal of Empirical Legal Studies* 4: 835–60.
- Michel, P., J. L. Quenon, A. M. de Sarasqueta, and O. Scemama. 2004. "Comparison of Three Methods for Estimating Rates of Adverse Events and Rates of Preventable Adverse Events in Acute Care Hospitals." *British Medical Journal* 328 (7433): 199–204.
- Needleman, J., P. Buerhaus, M. Stewart, K. Zelevinsky, and S. Mattke. 2006. "Nurse Staffing in Hospitals: Is There a Business Case for Quality?" *Health Affairs* 25: 204–11.
- Ontario Ministry of Health and Long Term Care. 2006. "Ontario Guide to Case Costing" [accessed October 2, 2011]. Available at: <http://www.ontla.on.ca/library/repository/mon/22000/283320.pdf>.
- Pine, M., J. S. Harmon, A. Elixhauser, D. E. Fry, D. C. Hoaglin, B. Jones, R. Meimban, D. Warner, and J. Gonzales. 2010. "Enhancement of Claims Data to Improve Risk Adjustment of Hospital Mortality." *Journal of the American Medical Association* 297: 71–6.
- Preyra, C. 2004. "Coding Response to a Case-Mix Measurement System Based on Multiple Diagnoses." *Health Services Research* 39: 1027–45.
- Quan, H., G. Parsons, and W. Ghali. 2004. "Assessing Accuracy of Diagnosis-Type Indicators for Flagging Complications in Administrative Data." *Journal of Clinical Epidemiology* 57: 366–72.
- Raleigh, V., J. Cooper, S. Bremner, and S. Scobie. 2008. "Patient Safety Indicators for England from Hospital Administrative Data: Case-Control Analysis and Comparison with US Data." *British Medical Journal* 337: a1702.

- Rivard, P., S. Luther, C. Christiansen, Z. Shibe, S. Loveland, A. Elixhauser, P. Romano, and A. Rosen. 2008. "Using Patient Safety Indicators to Estimate the Impact of Potential Adverse Events on Outcomes." *Medical Care Research & Review* 65: 67–87.
- Romano, P., B. K. Chan, M. E. Schembri, and J. A. Rainwater. 2002. "Can Administrative Data Be Used to Compare Postoperative Complication Rates across Hospitals?" *Medical Care* 40: 856–67.
- Rosen, A., S. Zhao, P. Rivard, S. Loveland, M. Montez-Rath, A. Elixhauser, and P. Romano. 2006. "Tracking Rates of Patient Safety Indicators over Time: Lessons from the Veterans Administration." *Medical Care* 44: 850–61.
- Rosenthal, M. 2007. "Nonpayment for Performance? Medicare's New Reimbursement Rule." *New England Journal of Medicine* 357: 1573–5.
- Stuart, E. A., G. King, K. Imai, and D. E. Ho. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42 (8):
- Sutherland, J., and C. K. Botz. 2006. "The Effect of Misclassification Errors on Case Mix Measurement." *Health Policy* 79: 195–202.
- Tabak, Y. P., X. Sun, K. G. Derby, S. G. Kurtz, and R. S. Johannes. 2010. "Development and Validation of a Disease-Specific Risk Adjustment System Using Automated Clinical Data." *Health Services Research* 45 (6), Part I: 1815–35.
- Thomas, E. J., S. R. Lipsitz, D. M. Studdert, and T. A. Brennan. 2002. "The Reliability of Medical Record Review for Estimating Adverse Event Rates." *Annals of Internal Medicine* 136 (11): 812–6.
- Wachter, R., N. Foster, and R. Dudley. 2008. "Medicare's Decision to Withhold Payment for Hospital Errors: The Devil Is in the Details." *Joint Commission Journal on Quality & Patient Safety* 34: 116–23.
- Wahl, P. M., K. Rodgers, S. Schneeweiss, B. F. Gage, J. Butler, C. Wilmer, M. Nash, G. Esper, N. Gitlin, N. Osborn, L. J. Short, and R. L. Bohn. 2010. "Validation of Claims-Based Diagnostic and Procedure Codes for Cardiovascular and Gastrointestinal Serious Adverse Events in a Commercially-Insured Population." *Pharmacoepidemiology and Drug Safety* 19 (6): 596–603.
- Zhan, C., and M. R. Miller. 2003a. "Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization." *Journal of the American Medical Association* 290: 1868–74.
- . 2003b. "Administrative Data Based Patient Safety Research: A Critical Review." *Quality & Safety in Health Care* 12 (suppl 2): ii58–63.
- Zhan, C., B. Friedman, A. Mosso, and P. Pronovost. 2006. "Medicare Payment for Selected Adverse Events: Building the Business Case for Investing in Patient Safety." *Health Affairs* 25: 1386–93.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.