

# Identification of Incident Injuries in Hospital Discharge Registers

Rolf Gedeberg,<sup>a,b</sup> Henrik Engquist,<sup>a</sup> Lars Berglund,<sup>b</sup> and Karl Michaëlsson<sup>b,c</sup>

**Background:** Hospital discharge data on injuries constitute a potentially powerful data source for epidemiologic studies. However, reliable identification of incident injury admissions is necessary. The objective of this study was to develop a prediction model for identifying incident hospital admissions, based on variables derived from a hospital discharge register.

**Methods:** There were 743,022 hospital admissions for injury in Sweden 1998–2004. Of these, 23,920 were in the county of Uppsala and 24% of these people had previous injury admissions. To determine if these admissions were new injuries or readmissions for earlier injuries, we reviewed 817 randomly selected hospital records. A prediction model for incident injury admissions was developed on the basis of patient age, type of admission (urgent or elective), time interval from the previous injury admission, main diagnosis, and department type.

**Results:** The final prediction model showed good discrimination (c-statistic = 0.969). This model was applied to the validation dataset using the optimal cut-off level, and the resulting sensitivity and specificity were adjusted according to the proportion with a previous injury admission in each injury category. The injury with the highest proportion of possible readmissions was hip contusion (35%). Nevertheless, using the prediction model, incident hip contusions were identified with a sensitivity of 94% (95% confidence interval = 93%–95%) and a specificity of 95% (94%–97%). The accuracy was higher for all other injury categories.

**Conclusions:** Incident injury admissions can be accurately separated from readmissions using a prediction model based on information derived from hospital discharge data.

(*Epidemiology* 2008;19: 860–867)

Submitted 11 October 2007; accepted 19 February 2008; posted 8 July 2008. From the <sup>a</sup>Department of Surgical Sciences – Anaesthesiology and Intensive Care; <sup>b</sup>Uppsala Clinical Research Center; and <sup>c</sup>Department of Surgical Sciences – Orthopedics, Uppsala University Hospital, Uppsala, Sweden. Supported by The Laerdal Foundation for Acute Medicine (PD07/05); The Swedish Society of Medicine (200618295); The Swedish Research Council (K2006-73P-20267-01-4 and K2006-73X-13511-07-3).

Supplemental material for this article is available with the online version of the journal at [www.epidem.com](http://www.epidem.com); click on “Article Plus.”

Correspondence: Rolf Gedeberg, Uppsala Clinical Research Center, Uppsala University Hospital, SE-751 85 Uppsala, Sweden. E-mail: [rolf.gedeberg@sursci.uu.se](mailto:rolf.gedeberg@sursci.uu.se).

Copyright © 2008 by Lippincott Williams & Wilkins  
ISSN: 1044-3983/08/1906-0860  
DOI: 10.1097/EDE.0b013e318181319e

Hospital discharge data can be a potentially powerful data source for epidemiology. Hospital discharge registers have been shown to predict death in hospital after procedures such as coronary artery bypass grafting with similar discrimination as clinical databases.<sup>1</sup> Compared with a dedicated trauma register, hospital discharge data are a valid source for documenting the nature and severity of injuries.<sup>2</sup> The use of secondary data must, however, be accompanied by validation of data accuracy<sup>3</sup> and careful attention to methodologic problems that are specific for data of this type.<sup>4–6</sup>

Many of these obstacles can be overcome in a setting with unique personal identification numbers that enable accurate record linkage.<sup>7</sup> This allows identification of multiple hospital admissions of the same individual patient. An important remaining limitation is the difficulty in obtaining accurate estimates of incidence of a disease when the disease can be recurrent.<sup>8,9</sup> Injuries are a good example of this situation. Repeated hospital admissions for injury are common, and these can represent either admission for a new injury or readmission for a previous injury. Unfortunately, most hospital discharge registers lack a specific variable that identifies the first hospital admission for that particular injury.<sup>8</sup> Even when such a variable exists, its accuracy can be questionable. A specific variable identifying readmissions was less accurate (90%–95% correctly classified) compared with a probabilistic approach (98%–99% correctly classified).<sup>9</sup> A deterministic rule combining a unique person identifier with injury date was almost as accurate as the probabilistic approach but injury date was incorrect in 10% of the records in that particular study. Simply selecting only the first occurrence of injury in each individual<sup>10,11</sup> could bias estimates of both incidence and outcome. Recurrent injuries are not uncommon, and these patients may constitute a selection of the population with characteristics related to the outcome. The opposite strategy, to regard all injury admissions as incident, has been used<sup>12</sup> but could also potentially cause a bias of incidence estimates,<sup>8</sup> since readmissions to hospital after injury are common. One possible way to improve the identification of incident injury admissions is to combine information from other variables in the hospital discharge register to estimate the probability of the admission being incident. Possible predictors include age, sex, diagnoses, time intervals between admissions, type of admission, and type of department.

The primary objective of this study was to develop a prediction model, based on data available in the Swedish

hospital discharge register, to separate incident injury admissions from readmissions for a previous injury. A secondary objective was to elucidate the effect on the estimated incidence of hip fracture of using different definitions of incident hospital admissions. Hip fracture was selected as a model injury due to the high incidence with consequent high impact on health care consumption and because there is an ongoing discussion concerning the trend of hip fracture incidence and methodologic issues.<sup>13–16</sup>

## METHODS

### Study Population

All hospital admissions for injury during the period 1998–2004 were extracted from the Swedish Hospital Discharge Register. For these patients, all previous injury admissions were identified during the period 1993–2004 from the same register using the unique personal identification number and main diagnosis. This register is a complete national register maintained by the Swedish National Board of Health and Welfare, and has covered all inpatient care in Sweden since 1987. In addition to the unique personal identification number that is given to all Swedish citizens,<sup>7</sup> the register includes information about the main diagnosis (which should reflect the main reason for hospital admission), comorbidity (up to 7 secondary diagnoses), surgical procedures, type of department, and type of admission (whether elective or urgent). There is no variable to indicate whether the admission is the first (incident or index) admission for an injury or a readmission. Since 1997 the diagnoses have been coded according to the International Classification of Diseases version 10 (ICD-10).<sup>17</sup>

We identified 743,022 hospital admissions, of which 23,920 were treated primarily in 1 of the 2 hospitals in the county of Uppsala. Uppsala University Hospital is an 1100-bed tertiary care facility and Enköping Hospital is a small local hospital. Together they serve a population of 302,564 (as of 31 December 2004). Direct transfers from hospitals outside the county of Uppsala were identified using record linkage and were excluded from the analysis. A weighted random sample was drawn from the injury admissions to these 2 hospitals by combining 2 simple random samples: (a) 10% of all admissions where the patient had at least 1 previous injury admission within the past 5 years and (b) 10% of all admissions where the patient had had a previous injury admission and where the time interval between the 2 admissions (from the date of discharge to the next admission date) was less than or equal to 360 days. There is no variable in the register that directly indicates whether the previous injury admission was for the same or a different injury episode. The reason for oversampling short admission intervals was our postulation that the length of this interval is an important discriminator between incident injury admissions and readmissions for injury. From these random selections of a total

of 820 hospital admissions, 817 patient records could be retrieved. The study was approved by the regional Human Ethics Committee.

### Definitions

A hospital admission for injury was defined as having a main diagnosis from chapter XIX (S and T diagnoses) in ICD-10, with the exclusion of allergy (T78), complications of surgical and medical care (T80–T88 or external causes Y40–Y84 that includes adverse effects of therapeutic use of drugs), and late effects of injuries (T90–T98 or external causes Y85–Y89). All information in records corresponding to direct transfers to another department or hospital was retained and merged with the original injury admission record, allowing up to 80 ICD-10 diagnoses for each injury admission. Causes of injury are presented according to the External Cause of Injury Mortality Matrix, developed by the International Collaborative Effort on Injury Statistics.<sup>18,19</sup> Previous hospital admissions for injury were identified from the years 1993–2004, ie, during a time period of at least 5 years preceding each admission in the study cohort.

Based on the patient records, 2 of the authors (R.G. and H.E.) classified each admission as either an incident injury admission or a readmission, to create a reference outcome variable. For each admission the raters had access to person identification number, admission date, department, hospital and admission interval (time from the previous injury admission) extracted from the register. The admission was classified as incident if from the patient record it could be clearly established that the admission was the first for that particular injury event. When it was evident that there had been a previous hospitalization caused by the same injury event, then the admission was categorized as a readmission. After 8–10 months, a randomly selected subset (stratified according to rater) of 40 admissions was reclassified by a different rater, blinded to the previous classification. The interrater reliability was almost perfect with  $\kappa = 0.90$  (95% confidence interval [CI] = 0.75–1.00).

### Derivation and Validation of the Prediction Model

Potential explanatory variables were identified on the basis of clinical reasoning (Table 1). These variables were used in logistic regression modeling to develop a prediction model that could separate incident hospital admissions from readmissions.

Admissions with a diagnosis of poisoning or concussion were all found to be incident and the indicator variables for these diagnoses thus generated complete separation. These postulated predictor variables could not therefore be included in the logistic regression model. The number of admissions with poisoning was considered large enough to warrant a deterministic rule that all hospital admissions for poisonings could be considered as incident. The 77 admis-

**TABLE 1.** Potential Predictors, Derived From Variables in the Swedish Hospital Discharge Register, That Were Used in Logistic Regression Modeling to Separate Incident Injury Hospital Admissions From Readmissions

Variable	Type	Description
Sex	0/1	0 = female, 1 = male
Age	Continuous	
Elective admission	0/1	0 = urgent hospital admission, 1 = elective hospital admission
Same diagnosis	0/1	1 = the first 4 positions in the ICD-10 main diagnoses are identical for the present and the previous hospital admission
Admission interval	Continuous	The number of days that have elapsed from the discharge date of the previous injury admission to the admission date of the present injury admission
Rehab	0/1	1 = admission to a department for rehabilitation, geriatrics, physiotherapy, sports medicine, or daytime care
Poisoning	0/1	1 = main diagnosis T36–T65 or T96–T97 <sup>a</sup>
Concussion	0/1	1 = main diagnosis S060 <sup>a</sup>
Burn	0/1	1 = main diagnosis T20–T35 <sup>a</sup>

<sup>a</sup>According to ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th revision, World Health Organization, Geneva).

sions for poisoning were therefore excluded from the study population while admissions for concussion were included. From the remaining 740 admissions, 400 were randomly selected as a model derivation dataset and 340 were set aside as an independent dataset for validation of the final prediction model. The size of the model derivation dataset was chosen to have a sufficient size for model development and still small enough to leave a reasonably large validation dataset after data split.

All variables in Table 1 were tested for main effects. The size of the development dataset precluded testing all possible interactions, although 2 likely interactions, selected on the basis of our clinical experience, were tested. These were: same diagnosis  $\times$  admission interval and elective admission  $\times$  admission interval. The corresponding interactions for the squared admission interval were also tested. Another 2 interactions were tested in the later stages of the analysis process, namely age  $\times$  same diagnosis and age  $\times$  admission interval. Model reduction was attempted on the grounds of our clinical perception of the predictive strength of the variables, with the aim of maximizing the area under the receiver operating characteristics (ROC) curve (*c*-statistic) as a measure of discrimination. The least likely predictor was removed first, with assessment of the consequent change in the *c*-statistic. If there was no reduction of the *c*-statistic, the variable was removed from the model. The optimal cut-off for the probability that the admission was incident was determined from the ROC curve for the final model. This prediction model was then applied to the evaluation dataset,

using that cut-off level to identify incident admissions. The resulting sensitivity and specificity with 95% confidence interval (CI) were calculated. This reflects the performance of the model in a population where everyone has had a previous injury admission. In reality, the majority of patients (approximately 75%) have not had a previous injury admission and their admissions are therefore considered incident. The proportion of admissions for which there was a previous admission also varies depending on the type of injury. For each injury category, we therefore adjusted for the proportion of admissions without a previous injury admission ( $1 - w_{dx}$ ), using the equation  $S_{dx} = w_{dx} \times S_{model} + (1 - w_{dx})$ , where  $S_{model}$  is either sensitivity or specificity, as determined by applying the prediction model to the evaluation dataset,  $w_{dx}$  is the diagnosis-specific proportion of admissions for which there was a previous injury admission, and  $S_{dx}$  is the resulting diagnosis-specific sensitivity or specificity. The variance was adjusted accordingly, by multiplying the unadjusted variance by  $w_{dx}^2$ . The confidence interval for the difference between the proportions of misclassified hip fractures was derived with the percentile method from a bootstrap sample with 1000 replications.

The SAS version 9 (SAS Institute, Cary, NC) statistical package or the R statistical package<sup>20</sup> was used for all analyses.

## RESULTS

A comparison of baseline characteristics for the study cohort, the Uppsala county cohort and the entire national cohort of hospitalized injuries is presented in Table 2. In the national population, 22% of injury admissions had a previous injury admission and hence were possible readmissions. The corresponding proportion among the Uppsala county patients was 24%. In the study cohort, representing a random sample of possible readmissions in the county of Uppsala, 59% (481/817) were incident injury admissions and the remaining 41% (336/817) were readmissions. The proportion of incident injury admissions varied depending on the diagnosis (Table 3).

The time interval from the previous injury discharge date (admission interval) had a distinct nonnormal distribution (Fig. 1). After examining a logit plot of this variable, it was included in the logistic regression as a second-degree polynomial. Another notable characteristic of the reference dataset was that 35% of the readmissions did not have the same main diagnosis (4 positions in the ICD-10 code) as the previous injury admission. Even when looking at only the first 3 positions of the ICD-10 code, 26% of the readmissions had a different main diagnosis from their previous admission.

Indicator variables for poisoning and concussion could not be included in the model, since they gave complete separation. Interactions with squared admission interval and admission interval with age were not significant, did not improve the *c*-statistic, and were consequently excluded from the model. All other variables, except sex and burn, gave an

**TABLE 2.** Baseline Characteristics of Hospitalized Injuries in the Entire National Cohort, the Entire Uppsala County Cohort and the Study Cohort

	Swedish National Cohort (n = 743,022)	Uppsala County Cohort (n = 23,920)	Study Cohort (n = 817)
Admissions with a previous injury admission; %	22.3	23.6	100
Age (yrs); mean (SD)	52.2 (28.9)	51.2 (28.3)	53.3 (27.4)
Female; %	52.0	52.8	52.0
Elective admission; %	13.4	19.4	47.1
Injury diagnoses <sup>a</sup> ; %			
Head injury <sup>b</sup>	3.9	3.4	6.0
Concussion	13.1	10.1	4.8
Hip fracture	16.9	16.6	14.9
Poisoning	5.5	6.5	8.7
Thoracic	1.1	1.4	0.6
Abdominal	0.65	0.83	0.6
Cause of injury <sup>c</sup> ; %			
Cut/pierce	2.7	2.9	5.1
Motor vehicle traffic	6.7	7.9	7.2
Fall	64.8	62.0	55.3

Only selected injury diagnoses and causes of injury are presented.

<sup>a</sup>Injuries are categorized according to the Centers for Disease Control Injury Mortality Diagnosis Matrix<sup>24</sup> and comparisons are presented for some selected injury categories.

<sup>b</sup>Concussion, ICD-10 code S060, is excluded from this category.

<sup>c</sup>Causes of injury are categorized according to the International Collaborative Effort on Injury Statistics<sup>19</sup> and comparisons are presented for some selected causes of injury.

**TABLE 3.** Proportion of Incident Admissions as Determined by the Review of Patient Records in the Study Cohort (n = 817)

Diagnoses	No. Admissions	Incident Admissions <sup>a</sup> %
Intracranial injury <sup>b</sup>	44	39
Fracture of forearm	31	48
Fracture of lower leg/ankle	62	57
Fracture of shoulder and upper arm	30	57
Fracture of rib, sternum, thoracic spine	33	73
Fracture of femur	19	74
Fracture of hip	118	78
Fracture of lumbar spine and pelvis	39	82
Superficial injury of hip and thigh	19	90
Poisoning	77	100
Concussion	39	100

Only the most common main diagnoses in this cohort are presented.

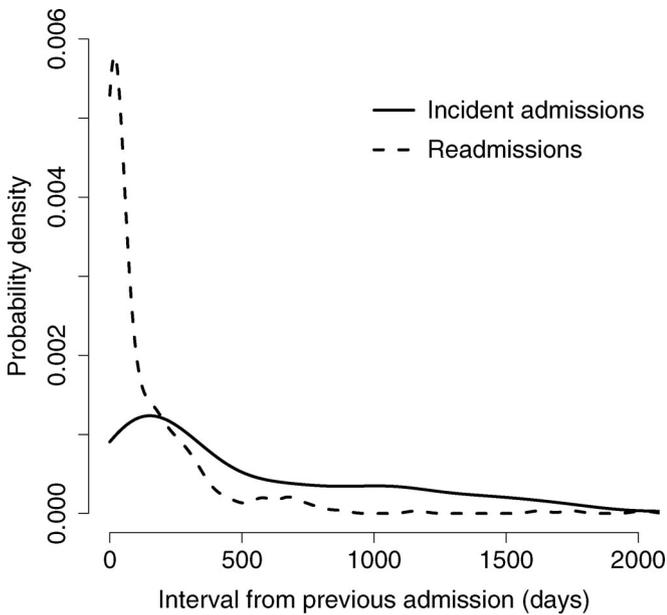
<sup>a</sup>Note that admissions with an interval from the previous injury admission of  $\leq 360$  days were oversampled to this cohort.

<sup>b</sup>Excluding concussion.

independent contribution to the predictive capacity as indicated by the c-statistic. The final model (eTable, available with the online version of the manuscript) had a c-statistic (corresponding to the area under ROC curve) of 0.969. The optimal cut-off level for the probability that the admission was incident was 0.4932, with a sensitivity of 93.7% and a specificity of 89.8%.

The prediction model was then applied to the validation dataset that consisted exclusively of possible readmissions. The resulting sensitivity was 83% (95% CI = 77%–88%) and specificity 87% (81%–92%). These measures of performance are conservative, since we can assume that all admissions without a previous injury admission are incident and these were not included in the analysis. The sensitivity and specificity therefore needed to be adjusted according to the proportion of possible readmissions. In the entire Uppsala county population of hospitalized injuries, 76% of the admissions had no previous injury admission within at least 5 years. Assuming that these admissions were incident and applying the prediction model only to the remaining 24% that were possible readmissions (admissions with a previous injury admission), the overall sensitivity increased to 97% (95%–97%) and the specificity to 97% (95%–98%). Since it is often of interest to study specific injury categories, and the proportion of suspected readmissions varies with the diagnosis, we also present sensitivity and specificity figures for injury categories, exemplified here by those with the highest proportion of possible readmissions (Table 4). The behavior of the prediction model (ie, the probability that the admission is incident) is also presented by illustrative combinations of predictors (Table 5).

When applying the prediction model to the dataset with all hospitalized injuries in Sweden, 96% (733,131/767,506) were identified as incident injuries. To illustrate the possible consequences of using different case definitions when esti-



**FIGURE 1.** Distribution of the time interval (admission interval) in days from the previous injury admission in the study population (n = 817).

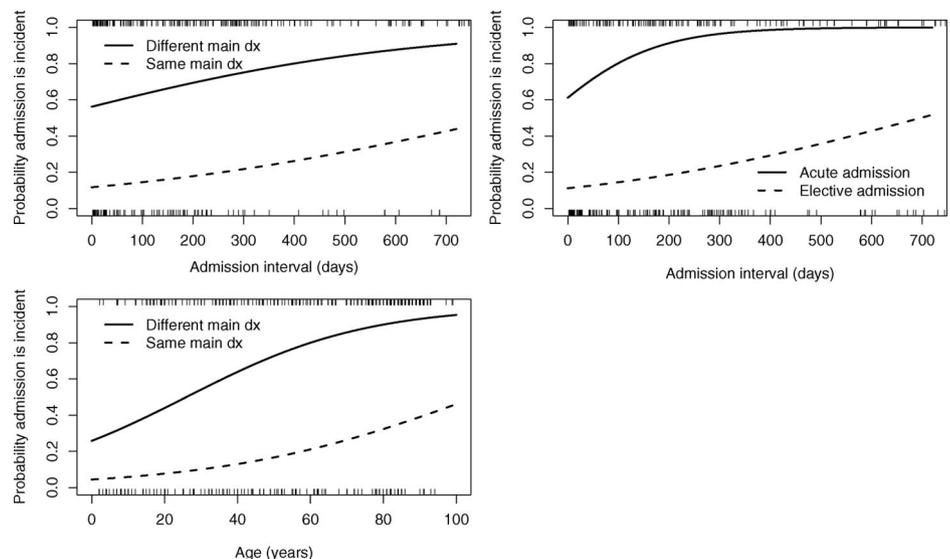
mating incidence from hospital discharge data we also applied the prediction model to calculate the incidence of hip fracture in the year 2004, using the Swedish population as of 1 November 2003 as the population at risk. There were 17,022 hospital admissions for hip fracture, of which 12,131 did not have a previous injury admission within at least 5 years. Using all admissions<sup>21</sup> to calculate the incidence proportion of hip fracture above 50 years of age resulted in an estimate of 504/100,000 population. If, on the other hand, only the individual's first admission for hip fracture during the reported calendar year was used for the calculation,<sup>14</sup> the

resulting incidence proportion was 489/100,000 population. When we finally applied the prediction model, 16,065 admissions were identified as incident, resulting in an incidence proportion estimate of 486/100,000 population. Incidence estimates stratified for age and sex are presented in Table 6. Although the latter 2 incidence estimates are equivalent, the method based on excluding multiple fractures within 1 calendar year will misclassify both incident admissions and readmissions to a greater extent. Even in our reference study population, where admission intervals shorter than 360 days were oversampled, 19% of readmissions for hip fracture had an admission interval exceeding 1 year. To illustrate the effect on classification of admissions, we applied the method excluding multiple fractures within 1 calendar year to the 122 patients with any mention of hip fracture in the study population. This method misclassified 28 (23%) of these admissions. Applying the prediction model instead to the same patients, resulted in a misclassification of 17 admissions (14%; 95% CI for the difference = 2%–16%).

### DISCUSSION

This study evaluates the performance of a prediction model for incident injury admissions, in a setting where unique personal identification numbers can be used to identify multiple admissions of the same patient and, among those, to eliminate transfers between departments and referrals between hospitals. Record linkage using unique personal identification numbers thus allows identification of a limited proportion of admissions where uncertainty remains as to whether they are readmissions for a previous injury or the first admission due to a new injury. When simple strategies are used to define incident cases, either by including all hospital admissions or only selecting first admissions, misclassification will inevitably occur. Our results indicate that

**FIGURE 2.** Probability that the hospital admission is incident visualized for the interactions incorporated in the prediction model: Same main diagnosis versus admission interval (upper left panel), elective admission versus admission interval (upper right panel), and same main diagnosis versus age (lower left panel). The distribution of data in each category is illustrated with the rug adjacent to each curve.



**TABLE 4.** Accuracy (Sensitivity and Specificity) of the Final Prediction Model for Separation of Incident Injury Admissions From Readmissions

Injury Region <sup>a</sup>	Nature of Injury <sup>a</sup>	No. Cases	Previous Injury Admission %	Accuracy of Prediction Model <sup>b</sup>	
				Sensitivity (95% CI)	Specificity (95% CI)
Hip	Contusion	448	35.3	94.0% (92.7%–95.2%)	95.3% (94.2%–96.5%)
Pelvis and lower back	Fracture	637	32.0	94.5% (93.5%–95.6%)	95.8% (94.8%–96.8%)
Traumatic brain injury	Fracture	551	28.1	95.2% (94.3%–96.1%)	96.3% (95.5%–97.1%)
Hip	Fracture	3,556	26.5	95.5% (94.7%–96.3%)	96.5% (95.7%–97.3%)
Vertebral column	Fracture	1,086	24.8	95.8% (95.0%–96.5%)	96.7% (96.0%–97.4%)
Upper extremity	Other specified injury	1,397	23.8	95.9% (95.2%–96.6%)	96.9% (96.2%–97.5%)
Upper extremity	Dislocation	441	23.4	96.0% (95.3%–96.7%)	96.9% (96.3%–97.5%)
Other lower extremity	Fracture	3,187	22.7	96.1% (95.5%–96.8%)	97.0% (96.4%–97.6%)
Upper extremity	Fracture	3,742	20.9	96.4% (95.9%–97.0%)	97.2% (96.7%–97.8%)
Thorax	Fracture	723	20.3	96.5% (96.0%–97.1%)	97.3% (96.8%–97.8%)
Other lower extremity	Other specified injury	831	20.2	96.5% (96.0%–97.1%)	97.3% (96.8%–97.8%)
Traumatic brain injury	Internal organ injury	3,202	18.8	96.8% (96.3%–97.3%)	97.5% (97.1%–98.0%)
Traumatic brain injury	Open wound	697	14.9	97.4% (97.1%–97.8%)	98.0% (97.7%–98.4%)
Total <sup>c</sup>		23,920	23.6	95.7% (94.4%–97.0%)	97.3% (95.8%–98.7%)

Results are presented for the injury categories with the highest proportion of admissions with a previous injury admission.

<sup>a</sup>Injury region and nature of injury are categorized according to the Centers for Disease Control Injury Mortality Diagnosis Matrix.<sup>24</sup>

<sup>b</sup>Sensitivity and specificity have been adjusted for the proportion of admissions with a previous injury admission within each injury category. Admissions without a previous injury admission are assumed to be incident.

<sup>c</sup>Note that this summation includes all injury categories, including those not presented in the table.

misclassification using these simple algorithms will range from minor to substantial. The degree of misclassification will depend on the chosen strategy to identify incident cases and what injury category is studied. The simple and inexpensive methodology described here allows researchers, administrators and policymakers the means to describe injury incidence with greater accuracy based on hospital discharge data, by making better use of all information available in the register. Adding new variables to a register is a much more

costly approach and not necessarily more accurate.<sup>9</sup> When we applied our prediction model based on variables available in the hospital discharge register, incident injuries could be identified with acceptable accuracy. Bias of the estimated regression coefficients due to model reduction should be limited, since most candidate predictors remained in the final model. In our example with hip fractures, we could demonstrate that, although a simpler method recommended in a recent methodologic paper<sup>14</sup> generated comparable incidence

**TABLE 5.** Predicted Probabilities That the Admission is Incident (the First Hospital Admission for That Particular Injury), Using the Final Prediction Model With Different Illustrative Values of Explanatory Variables (See Table 1)

Age (years)	Elective Admission	Rehabilitation	Same Diagnosis	Admission Interval (days)	Predicted Probability That the Admission is Incident
25	No	No	Yes	90	0.329
65	No	No	Yes	90	0.516
85	No	No	Yes	90	0.612
65	Yes	No	Yes	90	0.048
65	No	Yes	Yes	90	0.075
65	No	No	No	90	0.909
65	No	No	Yes	10	0.306
65	No	No	Yes	30	0.356
65	No	No	Yes	180	0.730
65	No	No	Yes	360	0.933
65	Yes	Yes	Yes	360	0.015

**TABLE 6.** Incidence Proportion of Hip Fracture<sup>a</sup> Per 100,000 Population<sup>b</sup> in the Year 2004

Age (years)	Sex	All Hospital Admissions	Prediction Model	Difference %
0–64	Women	19	18	5.1
	Men	21	20	5.2
65–69	Women	246	236	4.2
	Men	139	132	5.2
70–74	Women	542	524	3.5
	Men	248	237	4.9
75–79	Women	1,231	1,186	3.8
	Men	433	413	5.1
80–84	Women	2,779	2,667	4.2
	Men	751	732	2.6
85–89	Women	5,881	5,674	3.7
	Men	1,043	1,015	2.8
90–94	Women	10,708	10,333	3.6
	Men	1,236	1,203	2.7
95+	Women	4,314	4,216	2.3
	Men	4,009	3,973	0.9

Comparison of cases defined as all hospital admissions to cases defined by the prediction model.

<sup>a</sup>Defined as ICD-10 code S720–S722 as any mention of diagnosis.

<sup>b</sup>The Swedish population as of November 1, 2003 is used as the population risk.

estimates, this was because incident admissions and readmissions had been misclassified as the other about equally. Our prediction model is able to separate incident admissions from readmissions with greater accuracy.

Several variables in discharge data can be expected to predict whether the admission is incident, and this was confirmed in our study. In a recent innovative study, the length of the time interval between admissions was shown to be predictive of incident hip fracture admissions using a probabilistic method.<sup>4</sup> Some specific diagnoses can also be expected to be highly predictive in separating incident injuries from readmissions. In our study, a main diagnosis of poisoning or concussion was exclusively incident admission. The number of cases with poisoning in the study cohort was large enough to warrant a deterministic rule that all poisonings can be classified as incident. Concussion can probably be treated in a similar fashion, and this could increase the accuracy even further. However, before implementing such a deterministic rule for concussion, we consider it advisable to verify its accuracy in a larger population of patients. Identification of minor head injury in hospital discharge data can be inaccurate.<sup>2</sup> Another example of a diagnosis possibly highly predictive for readmissions is spinal cord injury. Such injury usually requires prolonged hospital care related to the same injury episode, and a new spinal cord injury will be highly unlikely. The incidence of spinal cord injury was too low, however, for this to be tested.

Having the same main diagnosis as the previous injury admission was also an important predictor. However, the

large proportion of readmissions with a main diagnosis different from that of the previous admission raises concern. Using a primary deterministic selection of cases with the same main diagnosis might introduce a substantial bias.

The type of admission, whether urgent or elective, is an obvious candidate predictor, and is available in many hospital discharge registers. In a recent extensive study examining injury-related hospital discharge data in 25 US states and the District of Columbia, all but 4 states reported whether the admission was elective or urgent.<sup>22</sup> Unfortunately, while an incident injury admission can be expected to be urgent in most cases, it is also conceivable that the first treatment of an injury can be managed in an outpatient clinic but that elective in-hospital care will later be required. It is also evident that a readmission to hospital for a previous injury can be elective or urgent.

There are other variables not present in Swedish hospital discharge data that could be of value in identifying incident injury admissions. A specific readmission indicator would obviously be helpful. However, in a previous validation study in a different setting, a readmission indicator had lower accuracy than a deterministic or probabilistic approach combining different variables.<sup>9</sup> Another variable that proved to be of value in that study was the date of injury, but this variable was also found to be missing or incorrect in 10% of the cases.

Our study has some potential limitations. The study cohort was relatively small, covering approximately 3% of the Swedish population. Although it was selected from a region (the county of Uppsala) it appeared to be representative of the national population. On the other hand, the main dataset is large and the study cohort represents a population-based selection. To verify the external validity of the prediction model, it should be tested on a larger hospital discharge dataset from a different setting. Differences in the predictor distribution between samples could alter the precision with which regression coefficients are estimated; this might affect the accuracy of the model.<sup>23</sup>

By applying the model on a separate dataset (data split) we were able to evaluate and quantify the performance of the model, enabling sensitivity analysis to illustrate the possible effect of misclassification when the model is applied in later studies. Although the number of hospital admissions used to develop the model is limited, the model appears statistically valid and the number of events in the validation dataset is adequate in relation to the number of variables evaluated. The method might to some extent generate an overly optimistic assessment of the predictive accuracy. Since the validation dataset was a simple random sample from the same population as the development dataset, this might bias the assessment of the predictive accuracy of the model in practice.<sup>23</sup>

Our results show the importance of accurate identification of incident hospital admissions. Readmissions are not

uncommon, and neither are new incident injury admissions when the patient has had a previous injury. This is further complicated by the fact that a large proportion of readmissions do not have the same diagnosis as the preceding admission, and that new injuries can occur within days after a preceding injury admission. Use of different strategies to handle (or disregard) the problem of possible readmissions can result in large variations in incidence estimates. Although methods based on a single predictor have been developed,<sup>4</sup> our study indicates that no single predictor is reliable on its own. A combination of predictors appears to be the most reliable way to achieve accurate identification of incident injury admissions. Combined with accurate record linkage using unique personal identifiers, this allows examination of the full extent of the treatment episode for a given injury.

### ACKNOWLEDGMENTS

We gratefully acknowledge valuable support in data management and SAS programming from Sören Gustafsson, Uppsala Clinical Research Center (UCR), in biostatistical methodology and R programming from Johan Lindbäck, biostatistician at UCR, and language revision of the manuscript by Maud Marsden.

### REFERENCES

1. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ*. 2007;334:1044.
2. McCarthy ML, Shore AD, Serpi T, et al. Comparison of Maryland hospital discharge and trauma registry data. *J Trauma*. 2005;58:154–161.
3. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25:435–442.
4. Sund R. Utilization of routinely collected administrative data in monitoring the incidence of aging dependent hip fracture. *Epidemiol Perspect Innov*. 2007;4:2.
5. Langley J, Stephenson S, Cryer C, et al. Traps for the unwary in estimating person based injury incidence using hospital discharge data. *Inj Prev*. 2002;8:332–337.
6. Cryer C, Langley JD, Stephenson SC, et al. Measure for measure: the quest for valid indicators of non-fatal injury incidence. *Public Health*. 2002;116:257–262.
7. Calltorp J, Adami HO, Astrom H, et al. Country profile: Sweden. *Lancet*. 1996;347:587–594.
8. Smith GS, Langlois JA, Buechner JS. Methodological issues in using hospital discharge data to determine the incidence of hospitalized injuries. *Am J Epidemiol*. 1991;134:1146–1158.
9. Alsop JC, Langley JD. Determining first admissions in a hospital discharge file via record linkage. *Methods Inf Med*. 1998;37:32–37.
10. Donohue JT, Clark DE, DeLorenzo MA. Long-term survival of Medicare patients with head injury. *J Trauma*. 2007;62:419–423.
11. Pentek M, Horvath C, Boncz I, et al. Epidemiology of osteoporosis related fractures in Hungary from the nationwide health insurance database, 1999–2003. *Osteoporos Int*. 2008;19:243–249.
12. Kannus P, Niemi S, Parkkari J, et al. Nationwide decline in incidence of hip fracture. *J Bone Miner Res*. 2006;21:1836–1838.
13. Boufous S, Finch C. Estimating the incidence of hospitalized injurious falls: impact of varying case definitions. *Inj Prev*. 2005;11:334–336 [erratum in: *Inj Prev*. 2006;12:135].
14. Brophy S, John G, Evans E, et al. Methodological issues in the identification of hip fractures using routine hospital data: a database study. *Osteoporos Int*. 2006;17:405–409.
15. Lonroos E, Kautiainen H, Karppi P, et al. Increased incidence of hip fractures. A population based-study in Finland. *Bone*. 2006;39:623–627.
16. Chevalley T, Guilley E, Herrmann FR, et al. Incidence of hip fracture over a 10-year period (1991–2000): reversal of a secular trend. *Bone*. 2007;40:1284–1289.
17. World Health Organization. *International Statistical Classification of Diseases and Health Related Problems ICD-10*. Geneva: World Health Organization; 2005.
18. Recommended framework for presenting injury mortality data. *MMWR Recomm Rep*. 1997;46(RR-14):1–30.
19. National Center for Health Statistics. International Collaborative Effort (ICE) on injury statistics: overview and implications. April 26, 2007. Available at: <http://www.cdc.gov/nchs/about/otheract/ice/matrix10.htm>. Accessed December 10, 2007.
20. R: A Language and Environment for Statistical Computing [computer program]. Version 2.5.1. R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2007.
21. Vestergaard P, Rejnmark L, Mosekilde L. Has mortality after a hip fracture increased? *J Am Geriatr Soc*. 2007;55:1720–1726.
22. Lawrence BA, Miller TR, Weiss HB, et al. Issues in using state hospital discharge data in injury control research and surveillance. *Accid Anal Prev*. 2007;39:319–325.
23. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.
24. Fingerhut LA, Warner M. The ICD-10 injury mortality diagnosis matrix. *Inj Prev*. 2006;12:24–29.