

Minimizing misclassification bias with a model to identify acetabular fractures using health administrative data

A cohort study

Andrew Adamczyk, MD, MSc^a, George Grammatopoulos, MBBS, DPhil^a,
Carl van Walraven, MD, MSc^{b,c,*}

Abstract

Acetabular fractures (AFs) are relatively uncommon thereby limiting their study. Analyses using population-based health administrative data can return erroneous results if case identification is inaccurate ('misclassification bias'). This study measured the impact of an AF prediction model based exclusively on administrative data upon misclassification bias.

We applied text analytical methods to all radiology reports over 11 years at a large, tertiary care teaching hospital to identify all AFs. Using clinically-based variable selection techniques, a logistic regression model was created.

We identified 728 AFs in 438,098 hospitalizations (15.1 cases/10,000 admissions). The International Classification of Disease, 10th revision (ICD-10) code for AF (S32.4) missed almost half of cases and misclassified more than a quarter (sensitivity 51.2%, positive predictive value 73.0%). The AF model was very accurate (optimism adjusted R^2 0.618, c-statistic 0.988, calibration slope 1.06). When model-based expected probabilities were used to determine AF status using bootstrap imputation methods, misclassification bias for AF prevalence and its association with other variables was much lower than with International Classification of Disease, 10th revision S32.4 (median [range] relative difference 1.0% [0%–9.0%] vs 18.0% [5.4%–75.0%]).

Lone administrative database diagnostic codes are inadequate to create AF cohorts. The probability of AF can be accurately determined using health administrative data. This probability can be used in bootstrap imputation methods to importantly reduce misclassification bias.

Abbreviations: AF = acetabular fracture, aOR = adjusted odds ratio, CT = computerized tomography, ICD = International Classification of Disease.

Keywords: acetabular fracture, health administrative data, misclassification bias, predictive models

1. Introduction

Although pelvic and acetabular fractures (AFs) account for only 1.5% of adult fractures, they are complex to treat.^[1] AF incidence ranges between 3 and 9.5 per 100,000 and typically follows a

bi-modal age distribution.^[2–4] Mechanisms of injury typically varies by age with high-energy trauma primarily responsible for AF in the young and while low-energy injuries, such as falling from a standing height, commonly cause AF in the elderly.^[2] AF incidence has quadrupled over the past 4 decades but seems to have recently stabilized.^[5]

Epidemiological studies focusing on AF are relatively uncommon (Table 1). These studies involve multiple countries including Finland,^[2] Scotland,^[3] France,^[6] and the United States.^[4,7] These data suggest that the preferred AF management in the young clearly consists of open reduction and internal fixation. In the elderly population, however, best management remains unclear due to the anatomical and medical implications involved in their treatment. AFs in the elderly have been of particular interest to trauma and arthroplasty surgeons alike since such elderly patients, compared to age matched patients with hip fracture, have longer hospital stays and higher risk-adjusted mortality.^[8] To date, outcome studies of the elderly AFs are limited to retrospective cohorts of limited size.^[9–11]

The lack of epidemiological studies focusing on management and outcomes of elderly AF patients hampers its study and advancement. This issue could be solved with population-based health administrative databases. Several population-based studies have created AF cohorts using diagnostic codes within hospitalization abstracts (Table 1). To our knowledge, however, the accuracy of diagnostic codes for AF has not been measured and no algorithms to identify patients with AF within health

Editor: Maya Saranathan.

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

^a Department of Surgery, University of Ottawa, Ottawa Hospital Research Institute, Canada, ^b Department of Medicine, University of Ottawa, Canada, ^c Department of Epidemiology & Community Medicine, University of Ottawa, Ottawa Hospital Research Institute, ICES, Canada.

* Correspondence: Carl van Walraven, ASB1-003, 1053 Carling Ave, Ottawa, ON K1Y 4E9, Canada (e-mail: carlv@ohri.ca).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Adamczyk A, Grammatopoulos G, van Walraven C. Minimizing misclassification bias with a model to identify acetabular fractures using health administrative data: a cohort study. *Medicine* 2021;100:52(e28223).

Received: 3 June 2021 / Received in final form: 29 October 2021 / Accepted: 24 November 2021

<http://dx.doi.org/10.1097/MD.0000000000002823>

Table 1
Description of published acetabular fracture cohort studies.

Study	Country	Case identification method	Sampling frame	Time period	N
Rinne ^[2]	Finland	ICD-10 S32.4 (1* or 2*)	All hospitals	1997–2014	5022
Laird ^[3]	Scotland	Trauma registry query	Single hospital	1998–2003	351
Melhem ^[6]	France	Not reported	All hospitals	2006–2016	32,614
Ferguson ^[7]	USA	Case registry all surgeries by single surgeon	Single hospital	1980–2007	1309
Best ^[4]	USA	ICD-9-CM 808.0 or 808.1	All hospitals	1990–2010	497,389

ICD = International Classification of Disease.

administrative data have been published or assessed. Measuring and optimizing the accuracy of AF identification using administrative data is essential to accurately study AF using population-based health administrative data with bias from case misclassification. In this study, we derived and internally validated a model using health administrative data which returns the probability that an AF was identified during a hospitalization. We then compared misclassification bias with AF identification using this model to that using single diagnostic codes for AF.

2. Methods

2.1. Study setting

The study took place at The Ottawa Hospital, a 1000-bed teaching hospital with 2 campuses that is the tertiary referral institution and trauma center in our region of approximately 1.3 million people. Annually, The Ottawa Hospital has more than 175,000 emergency department visits, 40,000 non-psychiatric admissions, and 50,000 surgical cases. The study was approved by the Ottawa Health Science Network Research Ethics Board (File: 20210026-01H).

2.2. Case identification

Our goal was to detect all AFs diagnosed at The Ottawa Hospital by reviewing the text reports of all computerized tomographic (CT) studies of the pelvis. To create this text search algorithm, we first collated a sample of AFs identified in our hospital's trauma registry. The trauma registry includes all patients who: presented to the hospital as a trauma code; were admitted under the trauma service; or had an injury severity score of 12 or higher upon presentation to the hospital. Detailed diagnostic, radiological, procedural, and outcome information is collected prospectively on each patient. For a separate analysis, we had retrieved the CT pelvis reports of all patients with radiographically confirmed AF in the trauma registry from January 2008 to December 2013 and January 2016 to December 2018. Using clinical experience and text analysis of this sample, we derived a text search algorithm that identified all CT pelvis reports indicating AF in this sample.

We then applied this text search algorithm to the reports of all CT pelvis studies at our hospital between January 1, 2008 and December 31, 2018. This time frame was chosen because of data availability. The reports of all CTs that were AF screen-positive were then manually reviewed (by AA) to identify patients with true AF. These people constituted all AF cases radiographically diagnosed at The Ottawa Hospital during the study period.

2.3. Creation of the AF model

We assumed that diagnostic and procedural codes would be important for a model that returned the probability of AF based

exclusively on administrative data. However, the number of distinct codes present in the discharge abstract of a large group of patients can be extensive. To identify diagnostic and procedural codes that might identify AF, we retrieved for the AF cases (identified in the previous step) all International Classification of Diseases, 10th revision (ICD-10) diagnostic codes and all Canadian Classification of Intervention procedural codes registered in their hospital discharge abstracts. Diagnostic and procedural codes were grouped by their first 3 and 5 alphanumeric characters, respectively, with the exception of the ICD-10 code specific to AF (S32.4). Code groups that were present in at least 5% of cases were independently ranked by 2 study members (AA, GG) regarding their clinical sensibility and likelihood to distinguish between patients with and without AF.

We then identified all non-psychiatric adult hospitalizations (defined as age exceeding 14) between January 1, 2008 and December 31, 2018 using our hospital's discharge abstract database. This dataset was linked to all AF cases to determine patients who were diagnosed with AF during their admission. We determined for each person the values of covariates from the hospital discharge abstract that we felt might be important to identify AF status including: age and sex; encounter urgency and ambulance status; hospitalization service; status of the diagnostic and procedural codes identified in the previous step; hospital length of stay; and death status.

Binomial logistic regression was then used to create a model that returned the probability of AF during each admission. To help prevent over-fitting, we limited the number of variables offered to the model to ensure no fewer than 20 degrees of freedom per AF case. To account for possible non-linear associations between continuous variables (i.e., age and hospital length of stay) and AF status, we used fractional polynomials with 2 terms identified with a transformation identification macro from Sauerbrei et al,^[12] thereby consuming a total of 4 degrees of freedom for each continuous variable. We did not use univariate inferential testing to screen for variable inclusion; instead, we ranked all variables by potential model relevance based on clinical experience and offered to the model all variables whose cumulative sum (degrees of freedom) was no less than one twentieth the total number of AF cases in our cohort. These variables were then all entered in the model with no subsequent variable selection based on their statistical association with AF status. This approach was used to minimize biased parameter estimates from variable selection using inferential testing ("testimation bias").^[13] We did not use other binary modelling methods, such as classification trees, random forests, neural networks, or support vector machines, because they do not necessarily improve model calibration.^[14]

Model performance was internally validated using optimism corrected c-statistics (for discrimination) and calibration in the large (for calibration) using methods described by Steyer-

berg^[13,15] using 200 bootstrap samples. We used Youden method to identify the expected AF probability having the greatest ability to discriminate between patients by AF status.^[16] We also measured the sensitivity of our model by using it to determine the expected AF probability of all AF patients from the trauma registry who were not used to derive our AF text search algorithm for pelvic CTs (i.e., all AF in the trauma registry admitted from January 1, 2014 to December 31, 2015).

We then quantified misclassification bias. First, we used the reference standard AF status to calculate true values of 9 statistics (prevalence of AF in study cohort; the association of AF with 2 continuous variables [age, hospital length of stay]; and the association of AF with 6 categorical variables [sex, admission urgency and ambulance status, packed red blood cell transfusions during the hospitalization, death status, and hospital procedure status]). With the exception of the latter (which was deemed present if patients had any Canadian Classification of Interventions code starting with '1' [indicating therapeutic intervention]), none of these statistics relied on administrative database codes and are very accurately measured.^[17]

We repeated measurement of these 9 statistics again after determining AF status using the ICD-10 code of S32.4 and the AF model. The latter approach used bootstrap imputation methods.^[18–20] Bootstrap imputation started by creating 1000 random bootstrap samples (with replacement) of the study cohort, each with a sample size identical to the original cohort. For each patient within each bootstrap sample, a uniformly distributed number between 0 and 1 was randomly selected; AF was imputed to be present if the random number was *below* the expected probability of AF for that patient (as determined from the AF model). Within each bootstrap sample, we measured all 9 statistics with the median value of all 1000 bootstrap samples used as the final point estimate and the 2.5th and 97.5th percentiles as the confidence intervals. We quantified misclassification bias as the unsigned relative difference in each of the 9 statistics compared to that achieved using true AF status. All analyses were conducted using SAS 9.4 (Cary, NC, USA).

3. Results

We initially identified 207 AF patients from our hospital's trauma registry. One hundred seventy-six of these patients (85.0%) had their acetabular imaging conducted at our hospital with CT reports available for analysis. All reports had at least 1 of the 4 text combinations that we found were important to identify AFs. This screen was then applied to the radiological reports of all pelvic CT studies at our hospital during the study period (n=296,588) and identified 1804 screen-positive reports conducted on 1558 patients. Manual review of these reports identified 1117 AFs in 908 patients. Of the latter, 728 unique patients were hospitalized when they were diagnosed with AF; the other cases had been transferred directly from our emergency department to another institution following their assessment at our center or were diagnosed based on ambulatory imaging.

During the 11-year study period, there were 438,098 non-psychiatric adult admissions to the hospital resulting in an AF incidence of 15.1 per 10,000 hospitalizations per year. AF admissions were distinct from other hospitalizations (Table 2). AF patients were slightly older and were much more likely to be male, arrive to the hospital by ambulance, or be admitted urgently. More than two-thirds of AFs were treated by the orthopedic or trauma service compared to only 10.1% of non-AF

admissions. The diagnostic code for AF (i.e., ICD-10 "S32.4") was present in only 51.2% of AFs and only 73.0% of patients with this code had an AF (i.e., S32.4 sensitivity and positive predictive value of 51.2% and 73.0%, respectively). A diagnostic code starting with "S32" was present in 33.9% of fractures but less than 1% of non-cases. Other relevant diagnostic codes most commonly present in AF related to co-injuries or injury mechanisms commonly found in AF. Similarly, the most common procedural codes identified in AF patients dealt with fixation of the pelvis or other loco-regional bones as well as local imaging studies. Hospital length of stay for AFs was much longer than the average but death risk was not distinctive.

Most of the selected covariates were significantly associated with AF status (Table 3). After adjustment for all model covariates, AF likelihood increased slightly with age but did not vary by sex. AF was notably more likely when patients were admitted urgently and under the orthopedics or trauma team. The strongest predictors for AF included codes for AF (adjusted odds ratio [aOR] 1782 [95% confidence interval 1289–2464]) and for lumber spine/pelvic fracture excluding AF (aOR 82.3 [62.2–108.9]). Fixation of the pelvis (aOR 2.93 [2.03–4.21]) and hip (aOR 4.75 [2.36–9.56]) were the procedural codes having the strongest association with AF. The optimism-corrected overall model fit (Nagelkerke R²: 0.618), discrimination (c-statistic: 0.988), and calibration (calibration slope: 1.06) was excellent. In patients without and with AF, the median (interquartile range; 5th–95th percentile) expected AF probability distribution was 0.01% (0.001%–0.03%; 0.00007%–0.008%) and 45.9% (6.2%–85.6%; 0.07%–98.5%), respectively. In 60 patients with AF from our trauma registry who were not used to generate our text search algorithm, the median (interquartile range; 5th–95th percentile) expected probability of AF was similar at 33.5% (7.1–71.3, 1.5%–99.4%).

Despite a very strong model, however, considerable misclassification occurred when we categorized the model-based expected AF probability (Table 4). We found that the most discriminating model-based expected AF probability was 0.09%. Using this cutpoint captured all but 44 of the 728 AF cases in the cohort (sensitivity 94.0%). However, only 3.7% of people with this expected AF probability or more actually had an AF (i.e., positive predictive value of 3.7%).

When AF status was determined using the AF model and bootstrap imputation, misclassification bias was always smaller – with 1 exception – when compared to that using the ICD-10 code S32.4 (Fig. 1). True AF incidence was 16.6 per 10,000 admissions; using the AF model and bootstrap imputation returned an identical value (Fig. 1A). In contrast, ICD-10 code of S32.4 significantly underestimated incidence by 29.5%. With the exception of "any procedure" (Fig. 1F), misclassification bias of association measures using the AF model and bootstrap imputation (median [range] unsigned relative difference 1.0% [0%–9.0%]) was always smaller than that when ICD-10 code S32.4 was used for case identification (median [range] unsigned relative difference 18.0% [5.4%–75.0%]). In most cases, estimates were biased towards the null; however, the association of AF using S32.4 with sex was significantly greater than true values (Fig. 1C).

4. Discussion

Population-based health administrative databases are very attractive for studying uncommon conditions like AF. However,

Table 2
Description of study cohort hospitalizations.

	Acetabular fracture		Total N = 438,098
	No N = 437,370	Yes N = 728	
Mean age (SD)	56.3 ± 20.7	60.8 ± 21.6	56.3 ± 20.7
Male	174,315 (39.9%)	421 (57.8%)	174,736 (39.9%)
Arrived by ambulance	124,439 (28.5%)	608 (83.5%)	125,047 (28.5%)
Admitted urgently	263,952 (60.3%)	720 (98.9%)	264,672 (60.4%)
Primary service			
Orthopedics	40,616 (9.3%)	311 (42.7%)	40,927 (9.3%)
Trauma	3652 (0.8%)	183 (25.1%)	3835 (0.9%)
Diagnostic codes (description)			
S32.4 (acetabular fracture)	138 (0.03%)	373 (51.2%)	511 (0.1%)
S32* (fracture of the lumbar spine and pelvis)	2604 (0.6%)	247 (33.9%)	2851 (0.6%)
S37 (injury of urinary and pelvic organs)	727 (0.2%)	53 (7.3%)	780 (0.2%)
S72 (fracture of femur)	8014 (1.8%)	80 (11.0%)	8094 (1.8%)
S22 (fracture of rib[s], sternum and thoracic spine)	3803 (0.9%)	157 (21.6%)	3960 (0.9%)
S27 (injury of other and unspecified intra-thoracic organs)	2174 (0.5%)	112 (15.4%)	2286 (0.5%)
V43 (car occupant injured in collision with vehicle)	835 (0.2%)	76 (10.4%)	911 (0.2%)
S82 (fracture of lower leg, including ankle)	4831 (1.1%)	90 (12.4%)	4921 (1.1%)
S36 (injury of intra-abdominal organs)	1818 (0.4%)	81 (11.1%)	1899 (0.4%)
Procedural codes (description)			
1SQ74 (pelvic fixation)	122 (0.0%)	208 (28.6%)	330 (0.1%)
30T20 (CT abdomen)	34,980 (8.0%)	181 (24.9%)	35,161 (8.0%)
1VA74 (hip fixation)	914 (0.2%)	41 (5.6%)	955 (0.2%)
1VA73 (hip joint reduction)	131 (0.0%)	38 (5.2%)	169 (0.0%)
1VA53 (implantation hip prosthesis)	10,827 (2.5%)	40 (5.5%)	10,867 (2.5%)
1VC74 (femoral fixation)	4501 (1.0%)	54 (7.4%)	4555 (1.0%)
3VZ20 (CT, MRI, or US of leg)	2116 (0.5%)	184 (25.3%)	2300 (0.5%)
Stay in days (SD)	7.2 ± 10.7	19.2 ± 16.1	7.2 ± 10.7
Any procedure done during admission	217,844 (49.8%)	481 (66.1%)	218,325 (49.8%)
Blood transfusion	43,139 (9.9%)	259 (35.6%)	43,398 (9.9%)
Patient died in hospital	17,941 (4.1%)	32 (4.4%)	17,973 (4.1%)

CT = computerized tomography, MRI = magnetic resonance imaging, US = ultrasound, SD = standard deviation.

* Excludes acetabular fractures (S32.4).

accurately identifying rare conditions in administrative data is always a challenge. In this study, we used our hospital's data warehouse to identify every AF case diagnosed at our hospital over 11 consecutive years. Using health administrative data, we found that the diagnostic code for AF had a sensitivity and positive predictive value of only 51% and 73%, respectively. Using this code to determine AF status underestimated AF prevalence by almost 30% and returned biased associations with other covariates. Using data found exclusively within health administrative databases, we created a very accurate model that returned the probability of AF for hospitalizations. When these expected probabilities were used to determine AF status using bootstrap imputation methods, misclassification bias was greatly reduced compared to that from using the ICD code for AF.

Our study makes several important points. **First**, our results highlight the potential misclassification resulting from using a single diagnostic code to identify AF. Although S32.4 (the ICD-10 code for AF) was very strongly associated with AF (with an aOR of 1783 [95% confidence interval 1289–2465]), the creation of an AF cohort using this code alone would miss almost half of cases and a quarter of this cohort would not actually have AF. In addition, we found that associations measured using this code for AF case identification frequently returned values that were importantly distinct from true values (Fig. 1). These results indicate the caution one must use when interpreting results from studies using non-validated codes for

case identification. **Second**, our model was well constructed using methods that addressed all of the key aspects highlighted in the PROBAST criteria^[21] for predictive model assessment. These include factors involving study participants (appropriate data sources and inclusion criteria), predictors (defined and available predictors), outcomes (determined appropriately and standardized, independent of predictors or model), and analysis (reasonable number of participants, appropriate handling of continuous variables, inclusion of all enrolled participants in the analysis, model predictors selected without univariate screening, and model performance measured adjusting for optimism). The AF model demonstrated exemplary optimism-adjusted performance explaining more than 66% of the observed variation in the cohort. It also had almost perfect discrimination (c-statistic 0.988) and was very well calibrated. Despite having such an accurate model, there was extensive misclassification when we categorized the model's expected AF probability for case identification (Table 3). This seemingly paradoxical result – a very accurate case-probability model returning misclassified disease status when a probability cutpoint is used – has been illustrated in other studies.^[18,20] These results highlight the need to use analytical methods, such as bootstrap imputation, that account for uncertainty of case ascertainment when using health administrative data. When case probability estimates from our AF model (Table 3) were applied using bootstrap imputation methods, prevalence estimates and measures of association with

Table 3**The acetabular fracture model.**

Variable	Parameter estimate (SE)	P value	Adjusted odd ratio (95% CI)
Intercept	-10.73 (0.52)	<.0001	-
Age increased by decade	0.066 (0.03)	.0195	1.07 (1.01, 1.13)
Male	0.004 (0.11)	.9718	1.00 (0.81, 1.24)
Arrived by ambulance	0.316 (0.14)	.0243	1.37 (1.04, 1.81)
Admitted urgently	2.775 (0.44)	<.0001	16.0 (6.80, 37.8)
Primary service: orthopedics	1.286 (0.13)	<.0001	3.62 (2.79, 4.69)
Trauma	0.729 (0.19)	.0001	2.07 (1.42, 3.02)
Presence of diagnostic codes			
S32 [‡] (acetabular fracture)	7.486 (0.17)	<.0001	1782.7 (1289.4, 2464.9)
S32 [*] (fracture of the lumbar spine and pelvis) [†]	4.411 (0.14)	<.0001	82.3 (62.2, 108.9)
S37 [*] (injury of urinary and pelvic organs)	0.586 (0.24)	.013	1.80 (1.13, 2.85)
S72 [*] (fracture of femur)	-0.444 (0.28)	.1164	0.64 (0.37, 1.12)
S22 [*] (fracture of rib[s], sternum and thoracic spine)	-0.179 (0.17)	.3012	0.84 (0.60, 1.17)
S27 [*] (injury of other unspecified intra-thoracic organs)	0.684 (0.21)	.0009	1.98 (1.32, 2.96)
V43 [*] (car occupant injured in collision with vehicle)	0.333 (0.21)	.1208	1.40 (0.92, 2.12)
S82 [*] (fracture of lower leg, including ankle)	0.016 (0.19)	.9337	1.02 (0.70, 1.47)
S36 [*] (injury of intra-abdominal organs)	0.398 (0.20)	.0508	1.49 (1.00, 2.22)
Presence of procedural codes			
1SQ74 (pelvic fixation)	1.074 (0.19)	<.0001	2.93 (2.03, 4.21)
3OT20 (CT abdomen)	0.520 (0.13)	<.0001	1.68 (1.30, 2.18)
1VA74 (hip fixation)	1.557 (0.36)	<.0001	4.75 (2.36, 9.56)
1VA73 (hip joint reduction)	2.031 (0.52)	<.0001	7.62 (2.76, 21.0)
1VA53 (implantation hip prosthesis)	0.560 (0.29)	.0571	1.75 (0.98, 3.12)
1VC74 (femoral fixation)	0.390 (0.31)	.2123	1.48 (0.80, 2.72)
3VZ20 (CT, MRI, or US of leg)	1.755 (0.17)	<.0001	5.78 (4.16, 8.03)
1/(length of stay+ 1) ²	1.592 (0.61)	.0096	-
1/(length of stay+ 1) ^{0.5}	-2.582 (0.54)	<.0001	-
Died in hospital	0.380 (0.23)	.098	1.46 (0.93, 2.29)

This table presents parameter estimates, *P* value, and adjusted odds ratio (with 95%) of all variables in the acetabular fracture model. Adjusted odds ratios exceeding 1 are associated with an *increased* likelihood of acetabular fracture. Odds ratios for length of stay are not presented because its association with AF is non-linear.

AF = acetabular fracture, CI = confidence interval, CT = computerized tomography, MRI = magnetic resonance imaging, US = ultrasound.

[†] Excludes acetabular fractures (S32.4).

key variables were very close to true values (Fig. 1). These results highlight the power of applying an accurate case-identification model using statistical methods that account for determination uncertainty. **Third**, it is commonly believed that misclassification will bias estimates towards the null. Our results indicate that this is not always the case (Fig. 1C).

Several issues should be kept in mind when assessing our results. **First**, our model has not been externally validated. This step will be important before it is applied to identify AF at a population-level. **Second**, our model can only be applied to health jurisdictions using other coding methods if ‘cross-walks’ are used to transform the codes used in our model to those native to the study center. Obviously, model accuracy should be reassessed if such steps are taken to confirm the validity of using this model. **Third**, it is likely that our AF case identification method will have missed some cases treated in our hospital during the study period. In our cohort of AF from the hospital’s trauma registry, we found that 15% of patients did not have any imaging done at the

hospital because imaging had been done at the referring hospital. When patients have no imaging done at our hospital, they will be missed by our case identification methods. However, the bias introduced into our model by this misclassification is unlikely to be extensive because of the overwhelming number of people in our cohort without AF.

In summary, we found that health administrative database diagnostic codes for AF are inadequate by themselves to create AF cohorts. We derived and internally validated a model that exclusively uses information available within health administrative database to return an accurate probability that AF is present during a particular hospitalization. When AF probability estimates were used to determine AF status using bootstrap imputation methods, misclassification bias was greatly reduced compared to that from using the ICD code for AF. If this model is validated in other centers, it could be used along with statistical methods accounting for its probabilistic nature to study AF at a population level.

Table 4**Operating characteristics of categorized expected acetabular fracture probability.**

	Acetabular fracture (N = 728)	No acetabular fracture (N = 437,370)
Expected AF probability ≥ 0.0009 (N = 18,436)	684	17,752
Expected AF probability < 0.0009 (N = 419,662)	44	419,618

We used Youden method to identify the most discriminative threshold for expected AF probability from the AF model (Table 3). This returned a sensitivity of 94.0% but positive predictive value of only 3.7%. AF = acetabular fracture.

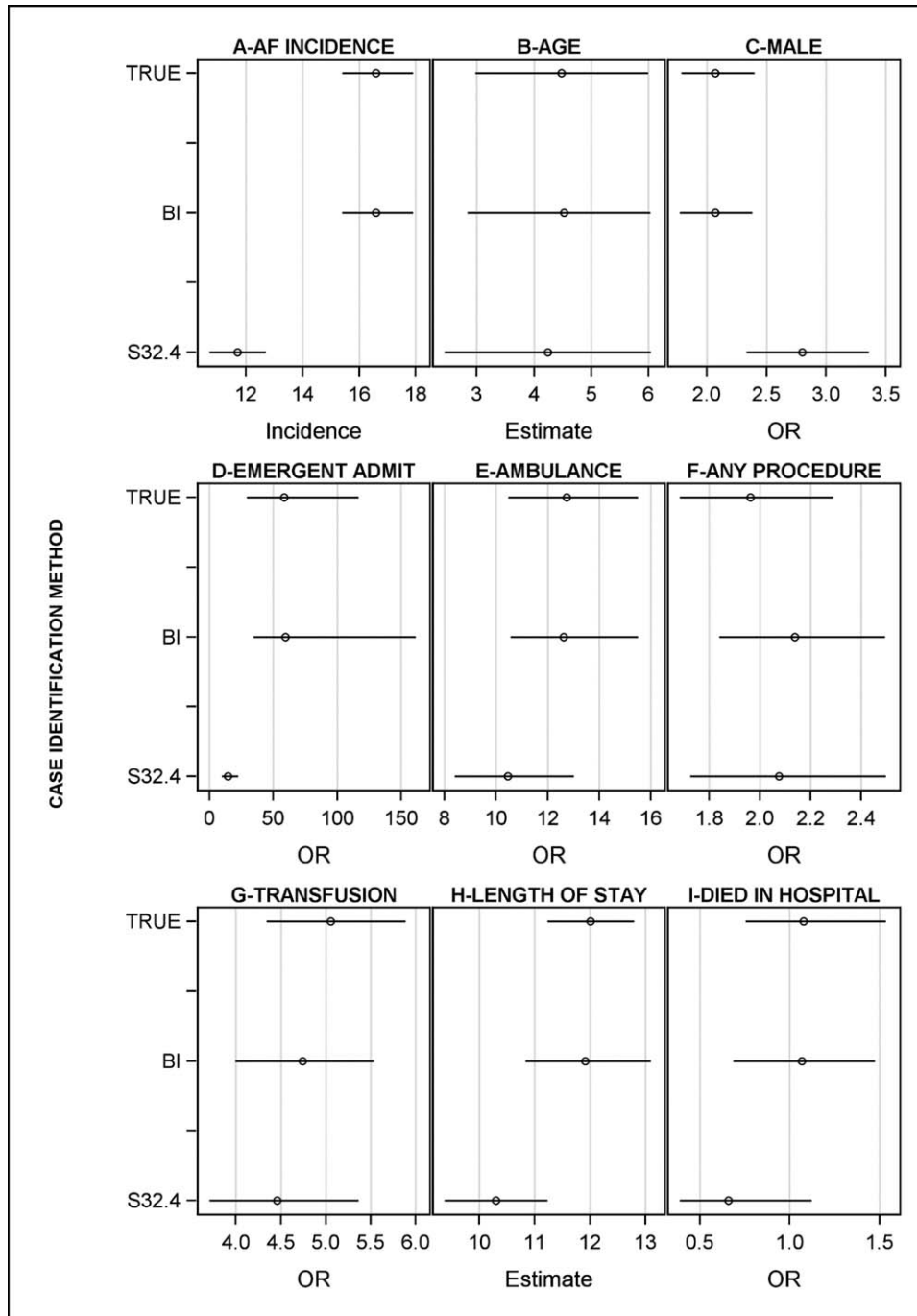


Figure 1. Misclassification bias when determining acetabular fracture status using the AF model or diagnostic code. This figure presents values for 9 statistics when AF status was determined with reference standard methods ("True"), with the AF model (Table 2) using bootstrap imputation ("BI"), or with the ICD-10 code for AF ("S32.4"). These statistics include AF incidence and the association of AF with continuous variables (AGE, LENGTH OF STAY) or binary variables (remaining variables). Associations are presented with 95% confidence intervals and were measured using linear regression for continuous variables (presented as the parameter estimate "Estimate") or logistic regression for binary variable (presented as the odds ratios "OR"). AF = acetabular fracture, ICD = International Classification of Disease.

Author contributions

Conceptualization: Carl van Walraven.

Data curation: Andrew Adamczyk, Carl van Walraven.

Formal analysis: Carl van Walraven.

Investigation: Carl van Walraven.

Methodology: Andrew Adamczyk, George Grammatopoulos, Carl van Walraven.

Project administration: Andrew Adamczyk, George Grammatopoulos, Carl van Walraven.

Resources: George Grammatopoulos, Carl van Walraven.

Software: Carl van Walraven.

Supervision: Carl van Walraven.

Validation: Andrew Adamczyk, Carl van Walraven.

Visualization: Carl van Walraven.

Writing – original draft: Carl van Walraven.

Writing – review & editing: Andrew Adamczyk, George Grammatopoulos.

References

- [1] Court-Brown C, Caesar B. Epidemiology of adult fractures: a review. *Injury* 2006;37:691–7.
- [2] Rinne PP, Laitinen MK, Huttunen T, Kannus P, Mattila VM. The incidence and trauma mechanisms of acetabular fractures: a nationwide study in Finland between 1997 and 2014. *Injury* 2017; 48:2157–61.
- [3] Laird A, Keating JF. Acetabular fractures: a 16-year prospective epidemiological study. *J Bone Joint Surg Br* 2005;87:969–73.
- [4] Best MJ, Buller LT, Quinlan SM. Analysis of incidence and outcome predictors for patients admitted to US hospitals with acetabular fractures from 1990 to 2010. *Am J Orthop (Belle Mead NJ)* 2018;47:
- [5] Kannus P, Parkkari J, Niemi S, Sievänen H. Low-trauma pelvic fractures in elderly Finns in 1970–2013. *Calcif Tissue Int* 2015;97:577–80.
- [6] Melhem E, Riouallon G, Habboubi K, Gabbas M, Jouffroy P. Epidemiology of pelvic and acetabular fractures in France. *Orthop Traumatol Surg Res* 2020;106:831–9.
- [7] Ferguson TA, Patel R, Bhandari M, Matta JM. Fractures of the acetabulum in patients aged 60 years and older: an epidemiological and radiological study. *J Bone Joint Surg Br* 2010;92:250–7.
- [8] Khoshbin A, Atrey A, Chaudhry H, et al. Mortality rate of geriatric acetabular fractures is high compared with hip fractures. A matched cohort study. *J Orthop Trauma* 2020;34:424–8.
- [9] Nicol GM, Sanders EB, Kim PR, Beaulé PE, Gofton WT, Grammatopoulos G. Outcomes of total hip arthroplasty after acetabular open reduction and internal fixation in the elderly-acute vs. delayed total hip arthroplasty. Does use of a quadrilateral surface plate improve outcome in elderly acetabular fractures? *J Arthroplasty* 2020; 36:605–11.
- [10] Nicol G, Sanders E, Liew A, et al. Does use of a quadrilateral surface plate improve outcome in elderly acetabular fractures? *J Clin Orthop Trauma* 2020;11:1045–52.
- [11] Sanders E, Dobransky J, Finless A, et al. Assessing outcome following open reduction internal fixation of acetabular fractures in the elderly. *Br Edit Soc Bone Joint Surg* 2021;103-B:
- [12] Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. *Comput Stat Data Anal* 2006;50:3464–85.
- [13] Steyerberg EW. Overfitting and optimism in prediction models. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2 ed New York: Springer; 2019. 95–112.
- [14] Christodoulou E, Ma J, Collins G, Steyerberg E, Verbakel J, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22.
- [15] Yoshida N, Hara M, Hasegawa M, et al. Reversible cochlear function with ANCA-associated vasculitis initially diagnosed by otologic symptoms. *Otol Neurotol* 2014;35:114–20.
- [16] Gonen M. *Single Continuous Predictor. Analyzing Receiver Operating Characteristic Curves with SAS*. Cary, NC: SAS Institute Inc.; 2007. 15–36.
- [17] Ontario Hospital Association, Ontario Ministry of Health, Hospital Medical Records Institute. Report of the Ontario Data Quality Reabstracting Study. 1991. Toronto, Ontario Hospital Association.
- [18] van Walraven C. Bootstrap imputation with a disease probability model minimizes bias from misclassification due to administrative database codes. *J Clin Epidemiol* 2017;84:114–20.
- [19] Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *Int J Biostat* 2011;7:6.
- [20] van Walraven C. Improved correction of misclassification bias with bootstrap imputation. *Med Care* 2018;56:e39–45.
- [21] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.