Title
Validation of a Rule-Based ICD-10-CM Algorithm to Detect Fall Injuries in Medicare Data

Journal
The Journals of Gerontology Series A, 79(7)

ISSN
1079-5006

Authors
Ganz, David A
Esserman, Denise
Latham, Nancy K
et al.

Publication Date
2024-07-01

DOI
10.1093/gerona/glae096

OXFORD JOURNALS The Journals of Gerontology Series A: Biological Sciences and Medical Sciences

J Gerontol A Biol Sci Med Sci. 2024 Apr 3;79(7):glae096. doi: 10.1093/gerona/glae096

Validation of a Rule-Based ICD-10-CM Algorithm to Detect Fall Injuries in Medicare Data

David A Ganz [1,2,✉], Denise Esserman [3], Nancy K Latham [4], Michael Kane [5], Lillian C Min [6], Thomas M Gill [7], David B Reuben [8], Peter Peduzzi [9], Erich J Greene [10]

Editor: Lewis A Lipsitz[11]

Author information Article notes Copyright and License information
PMCID: PMC11167485 PMID: 38566617

Abstract

Background

Diagnosis-code-based algorithms to identify fall injuries in Medicare data are useful for ascertaining outcomes in interventional and observational studies. However, these algorithms have not been validated against a fully external reference standard, in ICD-10-CM, or in Medicare Advantage (MA) data.

Methods

We linked self-reported fall injuries leading to medical attention (FIMA) from the Strategies to Reduce Injuries and Develop Confidence in Elders (STRIDE) trial (reference standard) to Medicare fee-for-service (FFS) and MA data from 2015–19. We measured the area under the receiver operating characteristic curve (AUC) based on sensitivity and specificity of a diagnosis-code-based algorithm against the reference standard for presence or absence of ≥1 FIMA within a specified window of dates, varying the window size to obtain points on the curve. We stratified results by source (FFS vs MA), trial arm (intervention vs control), and STRIDE's 10 participating health care systems.

Results

Both reference standard data and Medicare data were available for 4 941 (of 5 451) participants. The reference standard and algorithm identified 2 054 and 2 067 FIMA, respectively. The algorithm had 45% sensitivity (95% confidence interval [CI]: 43%–47%) and 99% specificity (95% CI: 99%–99%) to identify reference standard FIMA within the same calendar month. The AUC was 0.79 (95% CI: 0.78–0.81) and was similar by FFS or MA data source and by trial arm but showed variation among STRIDE health care systems (AUC range by health care system, 0.71 to 0.84).

Conclusions

An ICD-10-CM algorithm to identify fall injuries demonstrated acceptable performance against an external reference standard, in both MA and FFS data.

Keywords: Claims data, Encounter data, Fee-for-service Medicare, Medicare Advantage

---

Prevention of falls and consequent injuries in older adults has been a focus for clinicians and public health professionals, given the potentially serious consequences of fall injuries, including loss of ability to live independently. The first generation of research studies evaluating interventions to prevent falls and injuries relied heavily on self-reported data from participants as well as full-text medical record review (1), which when combined provided a high level of detail about fall and injury events. However, these data collection methods are subject to limitations including the cost of acquiring and reviewing data, failure of participants to recall their events (2), and participant loss to follow-up (3). Thus, there has been interest in using administrative data captured from the routine operations of health care systems to allow measurement of at least some types of fall events—those that lead to medical care and are thus potentially discoverable in administrative data. Administrative data have been used either on their own or adjunctively to measure fall injuries in pragmatic interventions based within the health care system (4–6) or in the community (7).

Diagnosis-code-based algorithms have been developed to identify falls and fall injuries leading to medical care in administrative data (8–11). However, a limitation of work to date has been the inability to validate the algorithms against a reference standard that is fully external (ie, from a data source that is independent of the one used by the algorithm). Using an administrative data source for both the reference standard and for the algorithm being validated could lead to bias if those who generate the codes in the administrative data (eg, health care providers and professional coders) select both the codes that count toward the reference standard and those that are used by the algorithm to predict the reference standard, thus inducing a correlation based on individual coding practices.

Where validation exists, it has been done in the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), which is no longer in use in the United States; and validation was limited to fee-for-service (ie, traditional) Medicare recipients (9). The transition to ICD-10-CM, which took effect in the United States on October 1, 2015, markedly changed the total number of diagnosis codes available (from about 17 000 to over 155 000); much of this difference was due to a marked expansion in the specificity of coding for traumatic injuries (up from about 15% of all codes in ICD-9-CM to about 60% of all codes in ICD-10-CM) (12). In addition, fee-for-service Medicare data no longer represent the majority of Medicare beneficiaries; 51% of the eligible Medicare population was enrolled in Medicare Advantage in 2023, with this proportion projected to increase to 62 percent by 2033 (13). In recent years, the Centers for Medicare and Medicaid Services have been releasing encounter data from Medicare Advantage, allowing the possibility that, together with existing fee-for-service data, administrative data on the entire population covered by Medicare (over 95% of adults aged 65 and older) (14) are now available for research. Given this context, there is a need to validate algorithms developed for ICD-10-CM against a fully external reference standard and in both fee-for-service Medicare and Medicare Advantage data.

In this analysis, we use data collected from the Strategies to Reduce Injuries and Develop Confidence in Elders (STRIDE) study to validate an existing ICD-10-CM rule-based algorithm (9,10) against the external reference standard of self-reported fall injuries leading to medical attention (FIMA). We also examine the performance of the algorithm in fee-for-service Medicare claims data and Medicare Advantage encounter data. Given the increased specificity of traumatic injury codes in ICD-10-CM, we hypothesized that the ICD-10-CM algorithm would demonstrate performance on par with or better than the prior ICD-9-CM algorithm.

Method

Data Source and Participants

STRIDE recruited 5 451 participants aged 70 and older at increased fall injury risk. Eighty-six primary care practices in 10 health care systems were randomized to either the STRIDE intervention (an individually tailored multifactorial fall injury prevention program led by a registered nurse fall care manager) or enhanced usual care. Details of STRIDE's design, screening and recruitment procedures, intervention, strategies for participant retention, protocol for outcome adjudication, and clinical outcomes have been described previously (5,15–21); STRIDE's protocol is available at clinicaltrials.gov (NCT02475850). Individuals were eligible for participation if they answered "yes" to at least 1 of 3 questions: (a) Have you fallen and hurt yourself in past year?, (b) Have you fallen ≥2 times in past year?, and (c) Are you afraid that you might fall because of balance or walking problems?

As part of STRIDE, health care systems provided information to link participants to Medicare data (19). We subsequently obtained fee-for-service claims and Medicare Advantage encounters for STRIDE participants covering STRIDE's follow-up period (2015–19) for this validation study. This study was approved by MassGeneral Brigham IRB, as an amendment to protocol 2015P000693. Reporting of this study complies with STARD requirements (see checklist in Supplementary Material) (22).

STRIDE participants whom we could successfully link to Medicare data constituted our study sample. See Supplementary Figure 1 for details regarding participant flow and exclusions. We restricted eligible follow-up time to start on October 1, 2015 (when ICD-10-CM was implemented in the United States), so that all events would be coded in ICD-10-CM.

Reference standard

The reference standard used in this analysis was self-reported FIMA. These data were collected via telephone interviews of participants (or their surrogates) every 4 months, over a median follow-up period of 2.3 years (interquartile range, 2.0–2.7 years) (21). Participants prospectively recorded their falls on fall calendars, which they could subsequently refer to as memory aids during the telephone interviews. If participants missed an interview, research staff asked about falls covering the time from the most recently completed interview to the current one. For any fall that resulted in an injury, participants were asked the date of the injury (which served as the index date for our analyses), details about the type of injury, whether the fall injury led to medical attention, and whether there was an overnight hospital stay. Fall injuries were considered FIMA if the participant indicated receiving medical attention or an overnight hospital stay. At each 4-monthly time point, completion of follow-up interviews (including surrogate death interviews) was greater than 93%; as a result, 71.8% of participants completed all follow-up interviews, and 9.2% completed all but 1 (18). Attrition rates from death or study withdrawal were 2.6 and 3.6 per 100 person-years of follow-up (PYF), respectively (18). Because FIMA are derived completely from self-report, they represent a fully external reference standard for the Medicare algorithm. We also delimited algorithm results to serious fall injuries and compared the delimited results against 2 alternative reference standards (see Supplementary Material).

Algorithm

Investigators have previously developed algorithms to detect serious fall injuries in administrative data (8,9). To validate an algorithm using ICD-9-CM diagnosis codes, Min and colleagues used cohorts from the nationally representative Health and Retirement Study with linked fee-for-service (Parts A and B) Medicare data (9). The algorithm used Medicare claims data from inpatient, outpatient, carrier, and skilled nursing facility files. The analysis used a composite reference standard consisting of survey data indicating a history of fall injury in the past 2 years that was serious enough to need medical care, combined with external cause of injury codes in Medicare data indicating falls (9). This algorithm has since been updated for use with ICD-10-CM codes (10), and this updated algorithm served as the starting point for our analyses. The algorithm has 3 variants ("acute care," "balanced," and "inclusive") that are progressively more sensitive (and less specific) at detecting events; in this analysis, we focus on the inclusive algorithm, as it demonstrated the highest overall sensitivity (results for the other 2 algorithms may be found in Supplementary Tables). Fall injuries are grouped by the algorithm into episodes of care; we used the start date of each episode as the index date for the fall injury (or injuries) and treated each episode as a single event. We made minor adaptations to the algorithm (see Supplemental Material) and used both Medicare fee-for-service claims files and the equivalent Medicare Advantage encounter files. For Medicare Advantage encounter data, we opted not to include additional data that health plans submitted from chart reviews, given that such information appears to be added non-randomly (23) and thus could bias our findings.

Statistical analysis

We first inspected the data for completeness and stratified data by source (fee-for-service Medicare or Medicare Advantage). We grouped data by Medicare file type (inpatient, skilled nursing facility, outpatient, carrier); by calendar year (based on reports of Medicare Advantage data becoming more complete over time) (24–26); by the 10 health care systems participating in STRIDE; and by STRIDE intervention and control groups. To determine whether there were systematic differences in groups with missing data, we compared baseline demographic and clinical characteristics across 4 groups: those with both reference standard and algorithm information available, those with only algorithm information available, those with only reference standard information available, and those with neither source of information. To understand the generalizability of STRIDE participants to the Medicare population aged 65 and older, we also compared the age, gender, race, and self-rated health of the STRIDE sample to published work using the Health and Retirement Study (27). We computed the total count of fall injury episodes detected by the algorithm and building on prior work (28), categorized these episodes by the most definitive injury type within the episode and whether an overnight hospitalization was present within the episode, as compared with the reference standards.

To characterize algorithm accuracy, we calculated the area under the receiver operating characteristic curve (AUC) using the trapezoidal rule (29), depicting sensitivity and specificity of algorithm-detected events against the reference standard, averaged across all available person-time in the data ("available" means that the reference standard information was present and Medicare data were similarly present for the time periods in question). We considered an AUC of 0.7 or greater to be acceptable (30).

To define algorithm sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), we partitioned follow-up time into nonoverlapping date windows that started on October 1, 2015 (for those who were enrolled in STRIDE on or before October 1, 2015) or the date of enrollment in STRIDE (for those who enrolled after October 1, 2015). A date window is a period of one or more consecutive days during which we analyzed both reference standard and algorithm data for the presence of at least 1 fall injury episode index date. Sensitivity is defined as the proportion of date windows containing one or more fall injury index dates from the reference standard that also contain one or more fall injury index dates detected by the algorithm. Specificity is the proportion of date windows containing no fall injury index dates from the reference standard that also contain no fall injury index dates detected by the algorithm. PPV is defined as the proportion of date windows containing one or more fall injury index dates detected by the algorithm that also contain one or more fall injury index dates from the reference standard. NPV is the proportion of date windows containing no fall injury index dates detected by the algorithm that also contain no fall injury index dates from the reference standard. Secondary measures included kappa statistics to better understand the degree of agreement beyond chance between the algorithm and the reference standard. Kappa statistics were computed directly from calculated sensitivity, specificity, and prevalence (31). We considered agreement (kappa) of <0 to be poor, 0.00–0.20 to be slight, 0.21–0.40 to be fair, 0.41–0.60 to be moderate, 0.61–0.80 to be substantial, and 0.81–1.00 to be almost perfect, in keeping with prior work (32).

To generate the receiver operating characteristic curve, we varied the length of the date windows from as short as 1 day to as long as 720 days, with each point on the curve representing a date window of specific length, over which the true positive rate and false positive rate were calculated. Narrow date windows are specific, in the sense that a reference standard event and an algorithm-detected event that fall within a tight date window are very likely to be referring to the same event, but at the cost of missing events where either the algorithm or the reference standard was in error concerning date (but correct with respect to event). In contrast, wide date windows are more likely to be sensitive (allowing room for error in date information), but at the cost of identifying some matches between the algorithm and reference standard that may be spurious. To allow delimiting of data to person-time in fee-for-service Medicare or Medicare Advantage, we also analyzed 1-month calendar date windows (eg, October 2015, November 2015, and so on) that aligned with monthly enrollment information in the Medicare Master Beneficiary Summary File.

For each of the reference standards, we also defined measures of sensitivity and specificity using the entire duration of each study participant's eligible follow-up period. We labeled the participant as "reference standard positive" if one or more FIMA were reported by the participant at one or more follow-up interviews; otherwise, we labeled the participant "reference standard negative." We labeled the participant as "algorithm positive" if one or more FIMA were detected by the algorithm over the entire eligible follow-up period; otherwise, we labeled the participant "algorithm negative."

To characterize the uncertainty around estimates and to account for the clustering of date windows within participants, we used the bootstrap technique to generate means and confidence limits for sensitivity, specificity, PPV, NPV, and kappa (33), resampling at the individual STRIDE participant level to generate 10 000 replicates, and then calculating 95% confidence intervals (CI) using the percentile method (34). In addition, to understand variation in algorithm performance, we stratified data by STRIDE's randomization arm (intervention vs control), by source of coverage (fee-for-service Medicare vs Medicare Advantage), and by each of the 10 participating health care systems. Depending on the analysis, we delimited data to Medicare fee-for-service versus Medicare Advantage person-time, or individuals continuously enrolled in Medicare fee-for-service versus Medicare Advantage. All analyses were run in SAS/STAT version 15.2, Cary, NC.

Results

Supplementary Figure 1 shows participant flow and reasons for exclusion from the final analytic sample. Supplementary Table 1 shows characteristics of the available Medicare fee-for-service and Medicare Advantage data. Data available per PYF for the outpatient facility file were notably greater in fee-for-service data than for Medicare Advantage but were similar for other files (inpatient, skilled nursing facility, and carrier). There was a marked variation in available data per PYF among the 10 health care systems participating in STRIDE.

Table 1 shows demographic and clinical characteristics of analyzed participants at time of enrollment. Of 5 451 STRIDE participants, 4 941 had both Medicare data and reference standard data available for analysis, of whom 3 104 (63%) had fee-for-service Medicare and 1 837 (37%) had Medicare Advantage. STRIDE participants with fee-for-service Medicare had lower proportions of individuals identifying as Hispanic as compared to participants with Medicare Advantage (4% vs 13%, respectively), higher proportions of individuals who had completed college (59% vs 43%, respectively), and lower proportions of those dually eligible for Medicare and Medicaid (4% vs 10%, respectively). Other characteristics appeared more balanced across fee-for-service and Medicare Advantage participants. Supplementary Table 2 shows baseline demographic and clinical characteristics of STRIDE participants by presence or absence of missing data. Although small sample sizes preclude drawing firm conclusions, those missing reference standard data only were notable for having higher rates of fall risk factors (use of mobility aid, and positive answers to screening questions for entry into the study) than the other groups. Supplementary Table 3 compares characteristics of STRIDE participants to a representative sample of Medicare enrollees aged 65 and older; STRIDE participants were older but healthier than the Medicare sample.

Table 1.

Baseline Demographic and Clinical Characteristics of STRIDE Participants with both Medicare Data and Reference Standard Data Available, by Medicare Status on the Month of study Enrollment

Characteristic	**Fee-for-service (*N* = 3 104)**	**Medicare Advantage (*N* = 1 837)**	**Total (*N* = 4 941)**
Age, y	79.6 ± 5.9	79.6 ± 6.5	79.6 ± 6.2
Women, *N* (%)[*]	1 865 (60.1)	1 182 (64.3)	3 047 (61.7)
Race, *N* (%)[*]			
White	2 841 (91.5)	1 666 (90.7)	4 507 (91.2)
Black	180 (5.8)	84 (4.6)	264 (5.3)
Other or unknown	83 (2.7)	87 (4.7)	170 (3.4)
Hispanic ethnic group, *N* (%)[*]	126 (4.1)	238 (13.0)	364 (7.4)
Educational level, *N* (%)			
High school graduate or less	555 (17.9)	559 (30.4)	1 114 (22.5)
Some college or equivalent	715 (23.0)	487 (26.5)	1 202 (24.3)
College graduate or higher	1 833 (59.1)	789 (43.0)	2 622 (53.1)
Unknown	1 (0.0)	2 (0.1)	3 (0.1)
Chronic conditions[†]			
No. per participant	2.1 ± 1.3	2.1 ± 1.4	2.1 ± 1.3
Fracture other than of the hip after 50 y of age, *N* (%)	1 083 (34.9)	558 (30.4)	1 641 (33.2)
Hip fracture after 50 y of age, *N* (%)	121 (3.9)	97 (5.3)	218 (4.4)

Characteristic	Fee-for-service (*N* = 3 104)	Medicare Advantage (*N* = 1 837)	Total (*N* = 4 941)
Clinically significant cognitive impairment, *N* (%)‡	2 (0.1)	6 (0.3)	8 (0.2)
Use of a mobility aid or inability to ambulate, *N* (%)	1 016 (32.7)	653 (35.5)	1 669 (33.8)
Screening risk questions for fall injuries, *N* (%)			
Fell 2 or more times in the past year	1 082 (34.9)	644 (35.1)	1 726 (34.9)
Had a fall-related injury in the past year	1 220 (39.3)	692 (37.7)	1 912 (38.7)
Was afraid of falling because of problems with walking or balance	2 702 (87.0)	1 548 (84.3)	4 250 (86.0)
Medicare coverage type, *N* (%)§			
Covered by Part A and Part B	2 964 (95.5)	1 837 (100.0)	4 801 (97.2)
Covered by Part A or Part B only	140 (4.5)	0 (0.0)	140 (2.8)
Dually eligible for Medicare and Medicaid, *N* (%)	117 (3.8)	188 (10.2)	305 (6.2)

Open in a new tab

Notes: Demographic and clinical characteristics shown here derive from the STRIDE study, except for age (date of birth), Medicare coverage type, and dual eligibility status, which are taken from the Medicare Master Beneficiary Summary File. Plus–minus values are means ± *SD*. Percentages may not total 100 because of rounding.

*Sex, race, and ethnic group were reported by the participant.

†Other chronic conditions reported included hypertension, cancer, arthritis, diabetes, chronic lung disease, myocardial infarction, stroke, congestive heart failure, and Parkinson's disease.

‡Based on 4 or more errors on the 6-item Callahan cognitive screening instrument or if the initial telephone interview was completed entirely by proxy.

[§]Medicare Part A covers mostly inpatient services; Medicare Part B covers mostly outpatient and professional services. Individuals with Medicare Advantage allow a Medicare Advantage health plan to administer their Medicare benefits.

Table 2 shows a count of fall injuries by injury type and hospitalization status, in the reference standard and detected by the algorithm. The total count of all FIMA was 2 054 for the reference standard and 2 067 for the algorithm. The algorithm detected fewer hospitalized events than the reference standard and a greater number of nonhospitalized events than the reference standard. The breakdown of events by injury type shows general concordance between the algorithm and the reference standard, with some exceptions (eg, fewer hospitalized events with a head injury, sprain or strain, bruising or swelling, or other serious injury detected by the algorithm than the reference standard).

Table 2.

Count of Fall Injuries by Injury Type and Hospitalization Status, in Reference Standards and in Algorithm, for Individuals with both Data Sources Available (*N* = 4 941)

	Hospitalized		**Not hospitalized**	
Injury type	**Self-reported**	**Algorithm**	**Self-reported**	**Algorithm**
Fracture	270	268	318	498
Dislocation	3	0	26	39
Cut with evidence of closure	32	27	188	227
Head injury	91	57	—	—
All other serious fall injuries*	126	40	—	—
Subtotal: All serious fall injuries	**522**	**392**	**532**	**764**
All other fall injuries leading to medical attention	35	39	965	872
Total: All fall injuries leading to medical attention	**557**	**431**	**1 497**	**1 636**

Open in a new tab

Notes: Data in the "self-reported" columns derive from the STRIDE study. Data in the "algorithm" columns derive from Medicare claims/encounters. Bolded values indicate results that are subtotals or totals.

*Overnight hospitalizations with sprain or strain, bruising or swelling, or other serious injury.

Table 3 shows algorithm performance at detecting reference standard events in the same calendar month. The algorithm had 45% sensitivity (95% CI: 43%–47%) and 99% specificity (95% CI: 99%–99%) compared to the reference standard; results for fee-for-service Medicare and Medicare Advantage were similar. These results translated into a PPV of 46% (95% CI: 44%–49%) and a

NPV of 99% (95% CI: 99%–99%), with a kappa statistic of 0.45 (95% CI: 0.43–0.47). Figure 1 shows receiver operating characteristic curves for the algorithm overall, and Table 4 shows underlying data. Figure 1 also shows results stratified by data source (Medicare Advantage vs fee-for-service), trial arm (intervention vs control), and by the 10 participating health care systems. The AUC was 0.79 (95% CI: 0.78–0.81) and did not differ substantively by Medicare fee-for-service vs Medicare Advantage data source or STRIDE trial arm, but showed variation by STRIDE health care system, with AUCs varying from 0.71 to 0.84 by health care system. Supplementary Tables 4 and 5 and Supplementary Figures 2 and 3 show algorithm performance for detecting serious fall injuries against the alternative reference standards. Delimiting algorithm results (and reference standards) to serious fall injuries resulted in higher AUCs and similar patterns of variation to the primary reference standard.

Table 3.

Algorithm Performance at Detecting Reference Standard Events in the Same Calendar Month

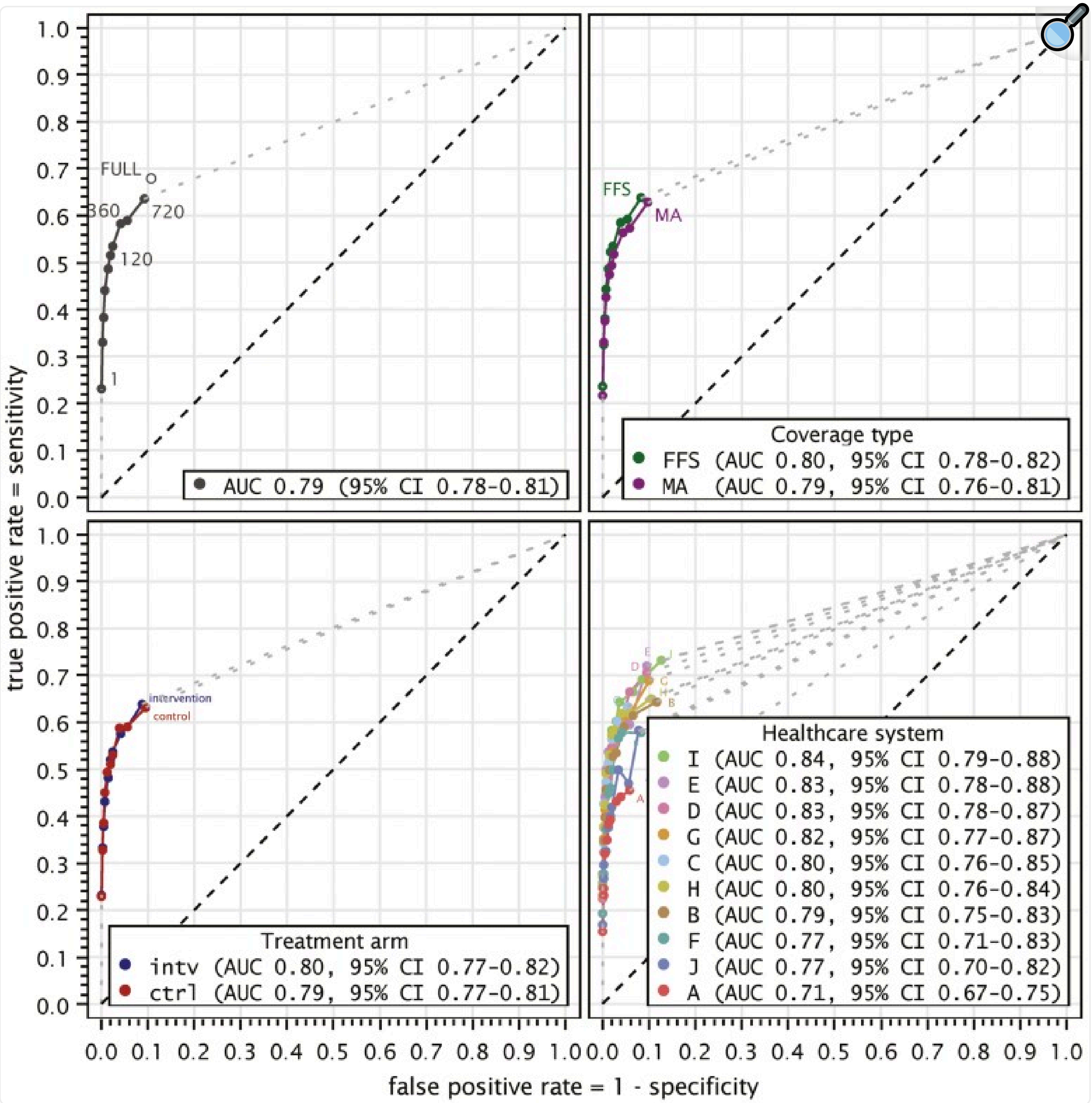
	Calendar months with ≥1 event in reference standard	Calendar months with ≥1 event detected by algorithm*	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Kappa (95% CI)
Overall	2 008	1 949	45% (43%–47%)	99% (99%–99%)	46% (44%–49%)	99% (99%–99%)	0.45 (0.43–0.47)
Fee-for-service Medicare data only	1 233	1 163	44% (41%–47%)	99% (99%–99%)	47% (44%–49%)	99% (99%–99%)	0.44 (0.42–0.47)
Medicare Advantage data only	775	765	45% (42%–49%)	99% (99%–99%)	46% (42%–50%)	99% (99%–99%)	0.45 (0.42–0.48)

Open in a new tab

Notes: CI = confidence interval.
Point estimates and confidence limits were generated from bootstrap resampling at the individual participant level to account for clustering of date windows within participants. Each accuracy measure was calculated within each replicate, and the mean and percentiles of each measure's bootstrap distribution provided its estimate and confidence interval.
*The overall count of events detected by the algorithm is greater than the sum of events detected in fee-for-service Medicare data only and Medicare Advantage data only because

some events were detected only via the combination of fee-for-service Medicare and Medicare Advantage data in individuals with both sources of data available.

Figure 1.



Open in a new tab

Area under the receiver operating characteristic curve for the Medicare data algorithm compared with the reference standard (self-reported fall injuries leading to medical attention). The y axis shows the true positive rate (sensitivity), and the x-axis shows the false positive rate (1-specificity). Inset boxes provide the area under the receiver operating characteristic curve and 95% confidence limits. The gray dashed lines demarcate the full receiver operating characteristic curve. The top left panel shows the overall results; numbers shown by each point represent the length of the date window in days over which the true positive rate and false positive rate were calculated, ranging from 1 day to 720 days. “FULL” represents a date window representing the full duration of follow-up for each participant. The top right panel shows results stratified by Medicare Advantage versus Medicare fee-for-service (among individuals continuously enrolled in Medicare Advantage or fee-for-service Medicare); The bottom left panel shows results stratified by intervention and control groups; and the bottom right panel shows results stratified by health care system. AUC = area under the receiver operating characteristic curve; CI = confidence interval; FFS = Medicare fee-for-service; MA = Medicare Advantage; intv = intervention; ctrl = control.

Table 4.

Algorithm Performance by Length of Date Window

Length of date window (in d)	Date windows with ≥1 event in reference standard	Date windows with ≥1 event detected by algorithm	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive value (95% CI)	Negative Predictive value (95% CI)	Kappa (95% CI)
1	2 036	2 067	23% (21%–25%)	100% (100%–100%)	23% (21%–25%)	100% (100%–100%)	0.23 (0.21–0.25)
7	2 028	2 003	33% (31%–35%)	100% (100%–100%)	33% (31%–36%)	100% (100%–100%)	0.33 (0.31–0.35)
15	2 007	1 957	38% (36%–40%)	100% (100%–100%)	39% (37%–41%)	100% (100%–100%)	0.38 (0.36–0.40)
30	1 954	1 907	44% (42%–46%)	99% (99%–99%)	45% (43%–47%)	99% (99%–99%)	0.44 (0.42–0.46)
60	1 898	1 838	49% (47%–51%)	99% (99%–99%)	50% (48%–53%)	99% (98%–99%)	0.48 (0.46–0.50)
90	1 826	1 772	51% (49%–54%)	98% (98%–98%)	53% (51%–55%)	98% (98%–98%)	0.50 (0.48–0.52)
120	1 831	1 764	53% (51%–56%)	98% (97%–98%)	55% (53%–58%)	97% (97%–98%)	0.52 (0.50–0.54)

Length of date window (in d)	Date windows with ≥1 event in reference standard	Date windows with ≥1 event detected by algorithm	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive value (95% CI)	Negative Predictive value (95% CI)	Kappa (95% CI)
240	1 580	1 514	58% (56%–61%)	96% (96%–96%)	61% (58%–63%)	95% (95%–96%)	0.55 (0.53–0.57)
360	1 408	1 322	59% (56%–62%)	94% (94%–95%)	63% (60%–66%)	93% (93%–94%)	0.55 (0.52–0.57)
720	1 035	951	64% (61%–67%)	91% (90%–92%)	69% (66%–72%)	89% (87%–90%)	0.56 (0.53–0.59)
Full follow-up time	1 473	1 376	68% (66%–70%)	89% (88%–90%)	73% (70%–75%)	87% (86%–88%)	0.58 (0.56–0.61)

Open in a new tab

Notes: CI = confidence interval.
A date window is a period of one or more consecutive days during which we analyzed both reference standard and algorithm data for the presence of at least one fall injury episode index date. Point estimates and confidence limits for all results were generated from bootstrap resampling at the individual participant level to account for clustering of date windows within participants. Each accuracy measure was calculated within each replicate, and the mean and percentiles of each measure's bootstrap distribution provided its estimate and confidence interval.

Discussion

For the reference standard of self-reported FIMA, we found an AUC for the ICD-10-CM algorithm of 0.79 (95% CI: 0.78–0.81). Although the AUC varied by STRIDE health care system, it remained at an acceptable level (≥0.7) at all sites. Our results are consistent with prior work by Min and colleagues, whose inclusive algorithm demonstrated a sensitivity of 68% and a PPV of 75% for matching events over a 2-year window, using a composite reference standard comprised of both external data (self-report) and internal data (external cause of injury codes from Medicare data) (9). Specificity and NPV were not reported. Our comparable analysis over a 720-day window showed a sensitivity of 64%, a specificity of 91%, a PPV of 69%, and an NPV of 89%. This overall similarity in algorithm performance may reflect counterbalancing effects: the increased specificity of ICD-10-CM codes may have boosted algorithm performance, while the rigor introduced by using a fully external reference standard may have offset these changes.

One unexpected finding was similarity of algorithm performance by source of data (fee-for-service Medicare vs Medicare Advantage), which was surprising given reports of incompleteness in Medicare Advantage data during the study time period (24,25). We surmise that performance looked similar between data sources due to the flexibility of the algorithm to detect qualifying events from either professional (eg, physician) or institutional (eg, hospital or emergency department) encounters, as well as the possibility that the participating health care systems in STRIDE were better-than-average reporters of data to Medicare Advantage health plans. The uniqueness of the 10 STRIDE health care systems (with respect to their willingness and ability to meet criteria for participation in a clinical trial) implies that caution is warranted in extrapolating results to Medicare Advantage data more broadly.

Another important finding was variation in algorithm performance among participating health care systems in STRIDE, which has implications for studies planning to use administrative data for outcome ascertainment. For clinical trials, randomizing units within a health care system as in STRIDE (rather than health care systems themselves) may help to avoid baseline imbalances in outcome detection due to differences in the algorithm's ability to detect events by health care system. For observational analyses, appropriately accounting for health care system differences in outcome detection will be important to avoid spurious findings.

There are limitations to our work. First, the reference standard used in this analysis cannot be considered a criterion standard, because even with the assistance of falls calendars, it is likely that older adults forget to report all falls or report the same event twice. Thus, the algorithm's true sensitivity and specificity may be better than what is observed. However, if the question is whether an administrative algorithm may be sufficiently accurate to replace self-reported data, then self-report is a reasonable reference standard. Second, the reference standard and administrative data derived from participants in a clinical trial who were selected to be at higher risk for fall injuries, and who were found to have a relatively high level of education; to the extent that these factors influenced the use of health care, results may not generalize to the national population. Third, the algorithm demonstrated high specificity but only moderate sensitivity; as a result, the algorithm will be better suited for comparative analyses (eg, comparing effectiveness, safety, or risk) than for measuring the total burden of fall injuries (35). Fourth, the algorithm appeared to systematically under-detect hospitalizations for injuries that are not always identified on imaging (head injury; bruising or swelling; and sprain or strain). This finding could relate to more variation in how these injuries were coded, or lack of systematic screening for injuries without imaging findings (eg, mild traumatic brain injury) (36).

Our work has several implications. The ability to apply a validated algorithm in Medicare Advantage as well as fee-for-service Medicare opens up new opportunities for research in Medicare Advantage participants, who are now a majority of Medicare beneficiaries (13). Second, the algorithm has maintained acceptable performance when applied to ICD-10-CM, which maintains the algorithm's relevance. Third, given that more than 95% of individuals aged 65 and older in the United States have Medicare coverage (14), both clinical trials and epidemiologic analyses could potentially use administrative data as a way to passively follow the older adult population for ascertainment of fall injuries. For this approach to be practical, however, Medicare Advantage data will need more timely releases, ideally on the same timeline as fee-for-service data (which are available on a quarterly basis). Fourth, the quality of administrative data coding still leaves significant room for improvement, likely because coding the cause of injury (eg, a fall) is not nationally mandated or linked to reimbursement (37); future policy initiatives could address this issue.

In conclusion, an ICD-10-CM algorithm to identify fall injuries in Medicare data demonstrated acceptable performance when compared with an external reference standard. This validation effort lays the groundwork for future interventions and observational studies to ascertain fall injury outcomes using routinely collected data available for the vast majority of older adults in the United States.

Supplementary Material

glae096_suppl_Supplementary_Figures_1-3_Tables_1-5

glae096_suppl_supplementary_figures_1-3_tables_1-5.zip (966.9KB, zip)

glae096_suppl_Supplementary_Data_S1

glae096_suppl_supplementary_data_s1.docx (166.6KB, docx)

glae096_suppl_Supplementary_Material

glae096_suppl_supplementary_material.docx (934.1KB, docx)

Acknowledgments

D.A.G., D.E., N.K.L., and E.J.G. conceptualized and designed this study. D.A.G., D.E., N.K.L., T.M.G., and D.B.R. were responsible for acquisition of data. All authors analyzed and interpreted the data. D.A.G. drafted the manuscript, and all authors provided critical revisions to the manuscript for important intellectual content. All authors approved the submitted manuscript.

Contributor Information

David A Ganz, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; Geriatric Research, Education and Clinical Center, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California, USA.

Denise Esserman, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.

Nancy K Latham, Boston Claude D. Pepper Older Americans Independence Center, Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

Michael Kane, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.

Lillian C Min, Division of Geriatric and Palliative Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor VA Medical Center, Center for Clinical Management Research and Geriatric Research Education Clinical Center (GRECC), Ann Arbor, Michigan, USA.

Thomas M Gill, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA.

David B Reuben, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA.

Peter Peduzzi, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.

Erich J Greene, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.

Lewis A Lipsitz, (Medical Sciences Section).

Funding

This work was supported by a grant from the National Institute on Aging at the National Institutes of Health (grant number R01AG071528); the STRIDE parent study was supported by the Patient-Centered Outcomes Research Institute with additional support from the National Institute on Aging (grant number U01AG048270); the Yale Pepper Center (grant number P30AG021342); and the Yale Clinical and Translational Science Award (grant number UL1TR001863).

Conflict of Interest

None.

Sponsor's Role

The organizations funding this study had no role in the design or conduct of the study; in the collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Department of Veterans Affairs, or the United States government.

References

1. Buchner DM, Hornbrook MC, Kutner NG, et al. Development of the common data base for the FICSIT trials. J Am Geriatr Soc. 1993;41(3):297–308. 10.1111/j.1532-5415.1993.tb06708.x [DOI] [PubMed] [Google Scholar]

2. Hannan MT, Gagnon MM, Aneja J, et al. Optimizing the tracking of falls in studies of older participants: Comparison of quarterly telephone recall with monthly falls calendars in the MOBILIZE Boston Study. Am J Epidemiol. 2010;171(9):1031–1036. 10.1093/aje/kwq024 [DOI] [PMC free article] [PubMed] [Google Scholar]

3. Griffin J, Lall R, Bruce J, Withers E, Finnegan S, Lamb SE; PreFIT Study Group. Comparison of alternative falls data collection methods in the Prevention of Falls Injury Trial (PreFIT). J Clin Epidemiol. 2019;106:32–40. 10.1016/j.jclinepi.2018.09.006 [DOI] [PubMed] [Google Scholar]

4. Ganz DA, Kim SB, Zingmond DS, et al. Effect of a falls quality improvement program on serious fall-related injuries. J Am Geriatr Soc. 2015;63(1):63–70. 10.1111/jgs.13154 [DOI] [PMC free article] [PubMed] [Google Scholar]

5. Bhasin S, Gill TM, Reuben DB, et al. ; STRIDE Trial Investigators. A Randomized Trial of a Multifactorial Strategy to prevent serious fall injuries. N Engl J Med. 2020;383(2):129–140. 10.1056/NEJMoa2002183 [DOI] [PMC free article] [PubMed] [Google Scholar]

6. Lamb SE, Bruce J, Hossain A, et al. ; Prevention of Fall Injury Trial Study Group. Screening and intervention to prevent falls and fractures in older people. N Engl J Med. 2020;383(19):1848–1859. 10.1056/NEJMoa2001500 [DOI] [PubMed] [Google Scholar]

7. Tinetti ME, Baker DI, King M, et al. Effect of dissemination of evidence in reducing injuries from falls. N Engl J Med. 2008;359(3):252–261. 10.1056/NEJMoa0801748 [DOI] [PMC free article] [PubMed] [Google Scholar]

8. Kim SB, Zingmond DS, Keeler EB, et al. Development of an algorithm to identify fall-related injuries and costs in Medicare data. Inj Epidemiol. 2016;3:1. 10.1186/s40621-015-0066-z [DOI] [PMC free article] [PubMed] [Google Scholar]

9. Min L, Tinetti M, Langa KM, Ha J, Alexander N, Hoffman GJ.. Measurement of fall injury with health care system data and assessment of inclusiveness and validity of measurement models. JAMA Netw Open. 2019;2(8):e199679. 10.1001/jamanetworkopen.2019.9679 [DOI] [PMC free article] [PubMed] [Google Scholar]

10. Hoffman G, Franco N, Perloff J, Lynn J, Okoye S, Min L.. Incidence of and County Variation in Fall Injuries in US Residents Aged 65 Years or Older, 2016-2019. JAMA Netw Open. 2022;5(2):e2148007. 10.1001/jamanetworkopen.2021.48007 [DOI] [PMC free article] [PubMed] [Google Scholar]

11. Mintz J, Duprey MS, Zullo AR, et al. Identification of fall-related injuries in nursing home residents using administrative claims data. J Gerontol A Biol Sci Med Sci. 2022;77(7):1421–1429. 10.1093/gerona/glab274 [DOI] [PMC free article] [PubMed] [Google Scholar]

12. Topaz M, Shafran-Topaz L, Bowles KH.. ICD-9 to ICD-10: evolution, revolution, and current debates in the United States. Perspect Health Inf Manag. 2013;10(Spring):1d. [PMC free article] [PubMed] [Google Scholar]

13. Ochieng N, Biniek JF, Freed M, Damico A, Neuman T.. Medicare Advantage in 2023: Enrollment Update and Key Trends. 2023; https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2023-enrollment-update-and-key-trends/ . Accessed September 21, 2023. [Google Scholar]

14. Bhaskar R, Noon J, O'Hara BJ.. The errors in reporting Medicare coverage: A comparison of survey data and administrative records. J Aging Health. 2019;31(10):1806–1829. 10.1177/0898264318797548 [DOI] [PubMed] [Google Scholar]

15. Bhasin S, Gill TM, Reuben DB, et al. Strategies to Reduce Injuries and Develop Confidence in Elders (STRIDE): A Cluster-Randomized Pragmatic Trial of a multifactorial fall injury prevention strategy: Design and methods. J Gerontol A Biol Sci Med Sci. 2018;73(8):1053–1061. 10.1093/gerona/glx190 [DOI] [PMC free article] [PubMed] [Google Scholar]

16. Gill TM, McGloin JM, Latham NK, et al. Screening, recruitment, and baseline characteristics for the Strategies to Reduce Injuries and Develop Confidence in Elders (STRIDE) Study. J Gerontol A Biol Sci Med Sci. 2018;73:1495–1501. 10.1093/gerona/gly076 [DOI] [PMC free article] [PubMed] [Google Scholar]

17. Reuben DB, Gazarian P, Alexander N, et al. The strategies to reduce injuries and develop confidence in elders intervention: Falls risk factor assessment and management, patient engagement, and nurse co-management. J Am Geriatr Soc. 2017;65(12):2733–2739. 10.1111/jgs.15121 [DOI] [PMC free article] [PubMed] [Google Scholar]

18. Gill TM, McGloin JM, Shelton A, et al. Optimizing Retention in a Pragmatic Trial of Community-Living Older Persons: The STRIDE Study. J Am Geriatr Soc. 2020;68(6):1242–1249. 10.1111/jgs.16356 [DOI] [PMC free article] [PubMed] [Google Scholar]

19. Ganz DA, Siu AL, Magaziner J, et al. ; STRIDE Investigators. Protocol for serious fall injury adjudication in the Strategies to Reduce Injuries and Develop Confidence in Elders (STRIDE) study. Inj Epidemiol. 2019;6:14. 10.1186/s40621-019-0190-2 [DOI] [PMC free article] [PubMed] [Google Scholar]

20. Gill TM, Bhasin S, Reuben DB, et al. Effect of a multifactorial fall injury prevention intervention on patient well-being: The STRIDE Study. J Am Geriatr Soc. 2021;69(1):173–179. 10.1111/jgs.16854 [DOI] [PMC free article] [PubMed] [Google Scholar]

21. Ganz DA, Yuan AH, Greene EJ, et al. Effect of the STRIDE fall injury prevention intervention on falls, fall injuries, and health-related quality of life. J Am Geriatr Soc. 2022;70(11):3221–3229. 10.1111/jgs.17964 [DOI] [PMC free article] [PubMed] [Google Scholar]

22. Bossuyt PM, Reitsma JB, Bruns DE, et al. ; STARD Group. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. 10.1136/bmj.h5527 [DOI] [PMC free article] [PubMed] [Google Scholar]

23. Meyers DJ, Trivedi AN.. Medicare advantage chart reviews are associated with billions in additional payments for some plans. Med Care. 2021;59(2):96–100. 10.1097/MLR.0000000000001412 [DOI] [PMC free article] [PubMed] [Google Scholar]

24. Medicare Payment Advisory Commission. Chapter 7: Ensuring the accuracy and completeness of Medicare Advantage encounter data. In: Report to the Congress: Medicare and the Healthcare Delivery System 2019; https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-

source/reports/jun19_ch7_medpac_reporttocongress_sec.pdf . Accessed December 30, 2022. [Google Scholar]

25. Medicare Payment Advisory Commission. Chapter 13: The Medicare Advantage Program: Status Report. Report to the Congress: Medicare Payment Policy 2020; https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/mar20_medpac_ch13_sec.pdf . Accessed December 30, 2022. [Google Scholar]

26. Serna L, Johnson A.. Medicare Advantage Encounter Data. 2022; https://www.medpac.gov/wp-content/uploads/2021/10/Encounter-data-MedPAC-01-Sept-2022.pdf . Accessed December 30, 2022. [Google Scholar]

27. Kim DH, Glynn RJ, Avorn J, et al. Validation of a claims-based frailty index against physical performance and adverse health outcomes in the Health and Retirement Study. J Gerontol A Biol Sci Med Sci. 2019;74(8):1271–1276. 10.1093/gerona/gly197 [DOI] [PMC free article] [PubMed] [Google Scholar]

28. Ganz DA, Gill TM, Reuben DB, et al. Costs of fall injuries in the STRIDE study: An economic evaluation of healthcare system heterogeneity and heterogeneity of treatment effect. Cost Effect Resource Allocation: C/E. 2023;21(1):49. 10.1186/s12962-023-00459-3 [DOI] [PMC free article] [PubMed] [Google Scholar]

29. Bamber D. Area above ordinal dominance graph and area below receiver operating characteristic graph. J Math Psychol. 1975;12(4):387–415. 10.1016/0022-2496(75)90001-2 [DOI] [Google Scholar]

30. Hosmer DW, Lemeshow S, Sturdivant RX.. Applied logistic regression. In: Balding DJ, Cressie NAC, Fitzmaurice GM, Johnstone IM, Molenberghs G, Scott DW, Smith AFM, Tsay RS, Weisberg S, eds. Wiley series in probability and statistics 398. Third edition. Wiley; 2013. 10.1002/9781118548387 [DOI] [Google Scholar]

31. Feuerman M, Miller AR.. The kappa statistic as a function of sensitivity and specificity. Int J Math Educ Sci. 2005;36(5):517–527. 10.1080/00207390500063967 [DOI] [Google Scholar]

32. Landis JR, Koch GG.. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174. 10.2307/2529310 [DOI] [PubMed] [Google Scholar]

33. Rutter CM. Bootstrap estimation of diagnostic accuracy with patient-clustered data. Acad Radiol. 2000;7(6):413–419. 10.1016/s1076-6332(00)80381-5 [DOI] [PubMed] [Google Scholar]

34. Margolis DJ, Bilker W, Boston R, Localio R, Berlin JA.. Statistical characteristics of area under the receiver operating characteristic curve for a simple prognostic model using traditional and bootstrapped approaches. J Clin Epidemiol. 2002;55(5):518–524. 10.1016/s0895-4356(01)00512-1 [DOI] [PubMed] [Google Scholar]

35. Wright NC, Daigle SG, Melton ME, Delzell ES, Balasubramanian A, Curtis JR.. The design and validation of a new algorithm to identify incident fractures in administrative claims data. J Bone Miner Res. 2019;34(10):1798–1807. 10.1002/jbmr.3807 [DOI] [PubMed] [Google Scholar]

36. Pozzato I, Meares S, Kifley A, et al. Challenges in the acute identification of mild traumatic brain injuries: Results from an emergency department surveillance study. BMJ Open. 2020;10(2):e034494. 10.1136/bmjopen-2019-034494 [DOI] [PMC free article] [PubMed] [Google Scholar]

37. Min J, Gurka KK, Kalesan B, Bian J, Prosperi M.. Injury burden in the United States: Accurate, reliable, and timely surveillance using electronic health care data. Am J Public Health. 2019;109(12):1702–1706. 10.2105/AJPH.2019.305306 [DOI] [PMC free article] [PubMed] [Google Scholar]

Associated Data

This section collects any data citations, data availability statements, or supplementary materials included in this article.

Supplementary Materials

glae096_suppl_Supplementary_Figures_1-3_Tables_1-5

glae096_suppl_supplementary_figures_1-3_tables_1-5.zip (966.9KB, zip) glae096_suppl_Supplementary_Data_S1

glae096_suppl_supplementary_data_s1.docx (166.6KB, docx)

glae096_suppl_Supplementary_Material

glae096_suppl_supplementary_material.docx (934.1KB, docx)
