HHS Public Access

Author manuscript

Med Care. Author manuscript; available in PMC 2023 November 01.

Published in final edited form as:

Med Care. 2022 November 01; 60(11): 852–859. doi:10.1097/MLR.0000000000001767.

Identifying Medicare Beneficiaries with Delirium

Lidia M.V.R. Moura, MD, PhD, MPH[a], **Sahar Zafar, MD, MBBS**[a], **Nicole M. Benson, MD, MBI**[b,c,d], **Natalia Festa, MD, MBA**[e], **Mary Price, MA**[c], **Maria A. Donahue, MD**[a], **Sharon-Lise Normand, PhD**[f,g], **Joseph P. Newhouse, PhD**[f,h,i,j], **Deborah Blacker, MD, ScD**[b,k], **John Hsu, MD, MBA, MSCE**[c,f]

[a]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.

[b]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

[c]Mongan Institute, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.

[d]McLean Hospital, Harvard Medical School, Belmont, Massachusetts.

[e]National Clinician Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA.

[f]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts.

[g]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

[h]Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

[i]Harvard Kennedy School, Cambridge, Massachusetts.

[j]National Bureau of Economic Research, Cambridge, Massachusetts.

Corresponding Author: Lidia M.V.R. Moura, Massachusetts General Hospital, 55 Fruit St., Boston, MA, 02114, lidia.moura@mgh.harvard.edu, Twitter: @LidiaMouraNVL.

Author contributions:

Dr. Moura contributed to the study design and conceptualization, data collection, data analysis, interpretation of data, and the original draft of the manuscript. Dr. Blacker contributed to study design and conceptualization; data collection, interpretation of data, and revising manuscript for intellectual content. Drs. Zafar, Benson, Donahue, and Festa contributed to data collection and revising manuscript for intellectual content. Mary Price, MA, and Dr Normand contributed to data analysis, interpretation of data, and revising manuscript for intellectual content. Dr. Newhouse contributed to study design and conceptualization, interpretation of data, and revising manuscript for intellectual content. Dr. Hsu obtained the funding and contributed to the study design, conceptualization, interpretation for the data, and revising the manuscript for intellectual content. **Statistical analysis** was conducted by Mary Price, MA, and Lidia Moura, MD, PhD, MPH, with supervision for statistical analysis by Sharon-Lise Normand, PhD.

Sponsor's Role:

This work was done as part of the fulfillment of Dr. Moura's doctoral degree in Population Health Sciences (Epidemiology) at the Harvard T.H. Chan School of Public Health.

Transparency Statement:

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Conflicts of Interest:

The authors declare no conflict of interest.

[k]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

Abstract

Background: Each year, thousands of older adults develop delirium, a serious, preventable condition. At present, there is no well-validated method to identify patients with delirium when using Medicare claims data or other large datasets. We developed and assessed the performance of classification algorithms based on longitudinal Medicare administrative data that included ICD-10 diagnostic codes.

Methods: Using a linked EHR-Medicare claims dataset, two neurologists and two psychiatrists performed a standardized review of EHR records between 2016–2018 for a stratified random sample of 1,002 patients among 40,690 eligible subjects. Reviewers adjudicated delirium status (reference standard) during this three-year window using a structured protocol. We calculated the probability that each patient had delirium as a function of classification algorithms based on longitudinal Medicare claims data. We compared the performance of various algorithms against the reference standard, computing calibration-in-the-large (CITL), calibration slope, and the area-under-receiver-operating-curve (AUROC) using 10-fold cross-validation (CV).

Results: Beneficiaries had a mean age of 75 years, were predominately female (59%), and non-Hispanic Whites (93%); a review of the EHR indicated that 6% of patients had delirium during the three years. While several classification algorithms performed well, a relatively simple model containing counts of delirium-related diagnoses combined with patient age, dementia status, and receipt of antipsychotic medications had the best overall performance (CV-CITL <0.001, CV-slope 0.94, and CV-AUC [0.88 95% CI: 0.84–0.91]).

Conclusions: A delirium classification model using Medicare administrative data and ICD-10 diagnosis codes can identify beneficiaries with delirium in large datasets.

Keywords

delirium; delirium prevalence; electronic health record; Medicare claims; validation

INTRODUCTION

Delirium is a preventable cognitive disorder characterized by an acute disturbance in attention and awareness and a change in cognition, such as confusion, disorientation, memory deficit, or language disturbance.[1–5] Older adults (≥ 65 years) are at high risk of developing delirium during an acute illness, as are individuals with an underlying neurocognitive disorder such as mild cognitive impairment or dementia.[1–4] Delirium is often described as an acute-onset neurological disorder,[6] however, delirium can go undetected and unidentified.[6,7] Delirium may lead to long-term cognitive decline,[8] increased mortality, healthcare complications, and hospital costs.[1–4,9]

Because delirium disproportionately affects older adults, its associated morbidity, mortality, and costs are projected to increase with population aging.[10] However, informed policy and

population health management are contingent upon the reliable ascertainment of disease status and healthcare utilization for persons with delirium.[11–13]

Understanding the prevalence of delirium and its relationship to national healthcare spending would be necessary to assess the effectiveness of public policies to prevent or reduce the delirium burden. Studies attempting to examine such questions generally require large administrative databases, e.g., Medicare administrative and insurance claims data, which would rely heavily on diagnosis codes. Currently, the accuracy of ICD-10 diagnosis codes for delirium is unknown. As with many other medical conditions, additional aspects of the diagnostic information, e.g., the diagnosis's setting or the frequency of diagnoses, could be informative.

The absence of a well-validated approach for classifying patients as having delirium or not when using large datasets has limited the knowledge base.[5,14,15] Estimates of new-onset delirium are highly variable (2% to 82%),[16–18] reflecting differences in definitions, classification methods, or data sources. Moreover, recent studies have found fewer patients with delirium diagnoses in claims data than clinically expected, which could reflect under-diagnosis, variable-diagnosis, and under-coding, among other challenges.[6,19]

The United States switched in October 2015 from the ICD-9 to the ICD-10 coding system; the system changed substantially between these versions (e.g., ICD-9-CM 293.0 Delirium due to conditions classified elsewhere and ICD-9-CM 293.1 Subacute delirium and under the ICD-10-CM system, F05- Delirium due to known physiological condition). No study has validated a classification algorithm for delirium using the ICD-10 codes.[5] Accurate classification of delirium status is a foundational step for national, population-based, or large dataset-based studies.

The purpose of our study was first to develop a reference standard of delirium based on clinical information contained in EHRs, then assess the performance of classification algorithms using longitudinal Medicare administrative data to classify delirium status among Medicare beneficiaries.

METHODS

Study design

This study of classification algorithms was designed to identify Medicare beneficiaries having at least one episode of delirium within a 3-year study timeframe. We used multiple sources of longitudinal data from 2016–2018, linked at the individual level, including EHR data with provider details, pharmacy, laboratory, and Medicare claims data with ICD-10 codes from an extensive healthcare delivery system, Mass General Brigham (MGB), which also served as a Medicare Accountable Care Organization (ACO). MGB includes two academic medical centers, seven community hospitals, three specialty institutions, and twenty-one community health centers. The dataset included all Medicare claims from Parts A, B, and D (i.e., hospital and physician services and prescription drugs) for subjects independent of the provider or location of care. During the study period, some nurses

within MGB collected Confusion Assessment Method (CAM) information on Neurology inpatients, but this use varied across patients and over time.[20,21]

Sample selection

We applied the following eligibility criteria using information assessed on 01/01/2016: 1) alive through the end of June 2016 (6+ months of observation data); 2) aged 65 years or older; 3) community-dwelling at the time of ACO entry; 4) enrolled in Medicare Parts A and B; and 5) enrolled for at least six months to the ACO during the observation period (01/01/2016 – 12/31/2018) or until death.[22] Overall, 40,690 beneficiaries met the eligibility criteria.

We then defined three mutually exclusive sampling groups (using only administrative data) for which we had three different a priori expectations of having disease: a) high-risk group, b) moderate-risk group, and c) a low-risk group. We separated the eligible group of subjects into one of the three sampling groups: A) patients with an ICD-10 diagnosis code suggestive of cognitive problems; B) patients with a clinic visit during the observation period with a neurologist, psychiatrist, geriatrician, neuropsychologist, or geriatric psychiatrist, but without an ICD-10 diagnosis suggestive of cognitive problems; and C) all other patients.[23]

Posteriorly, we used SAS to create a random number generator and sample without replacement of 2,100 subjects (out of a target population of 40,690 total eligible beneficiaries) within each of the three strata to create a review list and then randomly assigned patients on the list to one of the clinician reviewers (two neurologists, LM & SZ, two psychiatrists, NMB & DB). The patients were listed in random order within each list, and reviewers examined all available clinical data within the EHR during the observation period for each patient on their list, stopping once the total reviewed reached a pre-specified target of 1,000 subjects. Our reviewers examined the EHR data for 1,002 patients.

Among the 1,002 patients, 952 (95%) had available electronic health record data during the observation period; the remaining 5% had no medical visits to an outpatient clinic, emergency department, or hospital within the healthcare system during the three years. We dropped this 5% from the additional analyses. We also excluded 20 subjects (2%) who had insufficient clinical data in their records to determine whether they had delirium during the follow-up period (sensitivity analyses including the excluded subjects yielded comparable findings). There were 932 patients in the analytic sample who represented 40,690 patients in the eligible population. In the manuscript, we present our findings reweighted to the reference population of 40,600 was initially intended to increase privacy protection (Figure 1).

Clinical data review and disease adjudication

We developed a standard protocol for abstracting clinical data from the EHRs and for adjudicating delirium status based on this data (Supplementary Text). We based the adjudication protocol on the Diagnostic and Statistical Manual of Mental Disorders criteria, 5th Edition (DSM-V).[24]

Reviewers completed a three-month training period, during which we compared their ratings of a training sample of 100 subjects, iteratively refined the protocol, and assessed inter-rater reliability. All three reviewers examined the records for all subjects in the training sample in batches of 20 patients. After each batch, they discussed potential ambiguities within the protocol and refined the protocol to increase the accuracy of disease adjudication and agreement between reviewers (e.g., added a variable for diagnostic certainty and elaborated on the distinction between psychosis and delirium). Post-training interrater reliability was reasonable (kappa scores $\kappa \geq 0.80$ for adjudication of a patient's delirium status). After training, three reviewers (LM, SZ, and NMB) reviewed the full sample, while one reviewer served as supervisor (DB).

Reference standard: electronic health record review

Reviewers were blinded to sampling strata and claim information and classified each patient into one of the following categories: a) No evidence of delirium during the study time frame; b) At least one episode of delirium; or c) Unknown.

Because the reviewers relied on the available clinical information, potential evaluation errors or omissions in documentation could impact the information available for adjudicating delirium status. Accordingly, the reviewers rated their diagnostic certainty as high, moderate, or low, reflecting the level of confidence in assessing delirium status using the available EHR information. The reviewers also rated their confidence in each classification on a scale of "highly," "moderately," "mildly," or "not at all" confident. Clinical considerations informed reviewers' diagnostic certainty, data quality, and availability classifications.

When reviewers classified a patient as having "unknown" delirium status, they also recorded whether the classification resulted from inadequate documentation, conflicting reports on the same episode, or both. This real-world clinical standard is related to but distinct from a definition based on actual disease status, which reflects a combination of access to care, detection, and documentation. Our reference standard reflects the approach an expert clinician would use to classify a patient's delirium status based on the available clinical information.

Claims-based delirium diagnoses

We used the Medicare claims data from 2016-to 2018 to create delirium classification algorithms. For the "Base" algorithm, we used the ICD-10 diagnostic codes corresponding to a previously validated list of ICD-9 codes for delirium.[25] For other algorithms, "Refined," we created an indicator variable, incorporating a modified set of ICD-10 diagnostic codes determined to have strong face validity by clinical reviewers, based on our understanding of the DSM-V[24] and our clinical and research experience in the field. Therefore, we removed administrative codes inconsistent with current diagnostic criteria for delirium (e.g., ICD-9 code 293.84, "an anxiety disorder in conditions classified elsewhere" and ICD-10 code I674, "hypertensive encephalopathy" (Table S1).[23,26] Because information about the number of encounters and care settings in which patients received this diagnosis could be informative, we included this variable in several algorithms.

We also examined outpatient prescription medications received that could affect the risk of developing delirium or be used to manage it, such as receiving an antipsychotic drug (Table S2). We created categorical variables for having any drug dispensed and integer variables for counts of drug fill within the observation period. We obtained data on drugs dispensed from Medicare Part D files; Part D is a voluntary program, and an estimated 20% of our sample did not have Part D claims during the observation period. Finally, we included variables for age, sex, and dementia status in some algorithms because these factors may be associated with delirium risk.[22,27] We assessed dementia status using a validated method.[23]

Statistical analysis

We estimated a series of logistic regression models to assess the performance of each classification algorithm against the reference standard classification of delirium. We estimated 11 models with different combinations of delirium diagnosis codes and demographic variables (Table S3 and Supplementary Text). For example, one model (Model 8) used the "Refined" diagnosis codes, plus a count of hospitalizations (inpatient stays) and outpatient visits with delirium diagnoses, count of delirium-associated (antipsychotic) drug fills, and indicators for dementia status and age. Another model (Model 9) added sex to the Model 8 predictors.

The models estimated the probability of delirium for each subject during the observation period. Rather than selecting a specific probability threshold to determine the sensitivity and specificity of the models, we adopted an approach that incorporated the randomness of the outcome using a probability distribution. Specifically, for each model, we estimated the delirium status of each patient by sampling from a Bernoulli distribution with a probability equal to their likelihood of having delirium during the observation period as estimated from the model. We next compared the estimated delirium status with the reference standard and calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We repeated this process 1,000-times, utilizing Monte Carlo sampling to determine the mean, standard deviation (SD), minimum, and maximum estimates of the 1000 samples. Separately, we applied 10-fold cross-validation (CV) and plotted calibration curves and receiver operating characteristic (ROC) curves corresponding to each model. We reported calibration-in-the-large (CV-CITL; value of 0 is desirable), calibration slope (CV-slope; value of 1 is desirable), and the mean for the area under the ROC curve (CV-AUC), all with 95% confidence intervals. Because the probability of delirium increases with age and clinical practice patterns may differ across age groups, age effects could be non-linear; as a result, we repeated all analyses using categorical age variables (<75, 75–79, 80–84, 85+).

RESULTS

Population characteristics

Table 1 displays the characteristics of the reconstructed target population (n=40,690). The mean age of patients was 75 years, with females comprising 59% of the sample, and 5% had dementia. During the observation period, 58% had at least one emergency department visit or hospitalization, and 2% died.

Patients in Sample A had been assigned a diagnostic claim potentially consistent with cognitive problems. In contrast, patients in Sample B had a clinic visit during the observation period with a neurologist, psychiatrist, geriatrician, neuropsychologist, or geriatric psychiatrist, but without an ICD-10 diagnosis suggestive of cognitive problems, and patients in Sample C were all others not included in Samples A or B. Patients in Sample A were older, on average than patients in the other two samples, with mean ages of 81-years, 76-years, 74-years for Samples A, B, and C, respectively. Also, patients in sample A had overall worse outcomes than patients in the other two samples: The proportion of patients who had an ED visit or hospitalization within the 3-year observation period was 84%, 77%, and 48% for Samples A, B, and C, respectively; The 3-year mortality risk was 7%, 3%, and 2% for Samples A, B, and C, respectively.

Characteristics of patients with delirium

Table 2 displays the characteristics of the patients with (6%) and without (94%) delirium as adjudicated by clinician experts after reviewing the EHR. As expected, patients with delirium tend to be older, have higher baseline risk scores, and are more likely to have dementia than patients without delirium. Moreover, patients with delirium are more likely to have had an ED visit or hospitalization, received an antipsychotic drug, or died during the observation period than patients without delirium.

Delirium classification performance

Table S4 illustrates the performance characteristics of our preferred classification algorithm for delirium. In addition to its simplicity (it contained easily measurable inputs such as age, dementia status, and counts of hospitalizations or ED visits with a delirium diagnosis), the algorithm had the best calibration to the reference standard with CV-CITL of <0.001 and CV-slope of 0.94 (Figure 2) and yielded the best overall performance, with excellent discrimination (CV-AUC 0.88; 95% CI 0.84–0.91, Figure 3). The sensitivity was 47.2% (SD 2.7%; range 38.8–55.9%), specificity was 93.1% (SD 0.8%; range 90.3–95.4), positive predictive value was 57.3% (SD 3.1%; range 47.9–66.4%), and negative predictive value was 90.0% (SD 0.5%; range 88.7–91.6%). The supplement contains information on the performance of other models. Allowing age to be an ordered categorical instead of a continuous variable did not impact performance.

DISCUSSION

Diagnostic codes and large administrative datasets arguably represent the only feasible way to obtain national estimates of delirium within the entire Medicare program.[11,28] Counts of diagnostic codes with strong face validity, combined with information regarding context and demographic variables, can aid in identifying individuals with a high likelihood of true delirium. This type of validated approach is critical to the study of delirium within the Medicare program, which accounts for the majority of older adults in the United States. Accordingly, such an approach is essential to any research or policy utilizing large datasets to capture delirium status.[19]

One algorithm performed well and required only a few types of data, thus balancing both performance and simplicity; other algorithms performed nearly as well and represent viable options for some uses. Indeed, the optimal approach for identifying delirium in large datasets will depend on the question, design, and available data for a given study, as well as the performance of the classification algorithm. Similarly, the optimal use of the classification (e.g., individual delirium probabilities or dichotomous status indicators) may depend on a given study's specific needs.

This type of validated approach for identifying delirium using only administrative data arguably is much needed given the severity of delirium and its potential preventability.[29–31] The Medicare program currently incentivizes hospitals to reduce several inpatient complications such as catheter-related or nosocomial infections but has not yet introduced incentives concerning delirium, partly because of difficulty classifying beneficiaries with this condition.[32,33]

This study improves the classification of patients with delirium in several important ways. First, the preferred algorithm performs considerably better than previously published approaches.[25,34–36] For instance, a previous delirium classification algorithm by Kim et al. using an ICD-9 list alone had a low sensitivity (18%).[25] Improving the algorithm using antipsychotic use data had not improved sensitivity (30%).[25] Further, in a study by Inouye, the sensitivity of diagnosis by ICD-9 codes was only 3% compared to 74% sensitivity of a chart-based instrument.[35] Other studies have also reported underdiagnosed delirium when using ICD-9 codes in more restrictive or specific healthcare settings. For example, a study by Katznelson concluded that the hospital administrative database underestimates postoperative delirium rates after cardiac surgery.[34] One of our best-performing algorithms was more sensitive (47.2% [SD 2.7%]) when compared to previous studies and remained very specific (93.1% [SD 0.8%]).

Second, the preferred algorithm requires only Medicare administrative data and thus could be feasible to use with similar types of large datasets. Third, prior studies indicate that the ICD-9 diagnosis codes for delirium miss most cases.[25,37] The ICD-10 codes (implemented in 2015 in the United States) attempt to increase information granularity and improve disease classification relative to ICD-9 codes.[23,38] Our study is the first to assess the performance of an ICD-10-based diagnostic algorithm against a reference standard for delirium based on an expert clinician review of EHR data.

Fourth, the algorithm is both simple and transparent, which could be valuable for applications requiring clinical or stakeholder acceptance.

Finally, other studies examining the performance of delirium ascertainment methods were limited by small samples and by selection design such as restricting to a condition or setting, such as selection based on admission to intensive care unit stay, cardiac surgery, or admission to a skilled nursing facility.[25,34,37,39,40] In contrast, our stratified sampling approach using Accountable Care Organization enrollment (like a membership roster independent of condition or setting) arguably increases the generalizability of the findings.

Limitations and Future Directions

We are limited to information available within the EHRs, which included variable documentation of validated delirium screening assessments like the CAM. Data is contingent upon evaluations and documentation by clinicians, mainly in bedside encounters; thus, ascertainment is neither systematic nor random.[41] Not all staff from hospital floors documented observed delirium symptoms, calculated CAM scores or even screened high-risk patients (i.e., older adults with acute ischemic stroke).[13] Not all physicians screened for or documented cognitive changes in the medical record; in fact, hypoactive delirium cases could be frequently undiagnosed.[9] In our protocol, we did not count how many patients had CAM scores available because, in prior work, we have found that CAM use was sporadic. Instead, our reviewers attempted to adjudicate disease states based on all the available clinical data in the EHRs.

Among subjects adjudicated as having delirium, we observed at least one CAM score available in about a quarter of the issues. We validated delirium screening assessments like the CAM varied by institution, floor, department, care setting, and patient characteristics over time. We found the measurements sporadic, even in periods and feet that should have had routine CAM score assessments. This variable ascertainment in real-world practice is the topic of a separate line of work examining practice patterns and quality.

Our results might not be generalized to estimates using data from other hospital systems or other healthcare databases. All our subjects were aligned to a Medicare ACO based in a large academic health system in New England. While a little over half of the persons aged 65+ years nationally are enrolled in the fee-for-service Medicare program, only about a third of these are in ACOs.[42] The findings might not generalize to other ACOs, beneficiaries in standard fee-for-service Medicare, not in ACOs, or those enrolled in a Medicare Advantage plan.[43] It would be desirable to replicate our analysis in other samples.

Our approach also focused on establishing delirium status for each patient over the observation period rather than identifying episodes of delirium. We did not attempt to define the severity, etiology, or potential preventability of the condition. Moreover, because we did not identify specific episodes and allowed delirium status to change over time, we could not assess the temporal relationship between the antipsychotic medication fills and delirium episodes. These limitations suggest natural extensions to this initial work.

Additionally, 2% of the sample had limited clinical information, e.g., records limited to nonspecific clinical encounters such as phone call notes or focused preventive care-like routine eye exams. We suspect that low healthcare utilization among these patients suggests good health and cognition. Nonetheless, some of these individuals could have avoided presentation to care or, theoretically, had conditions that precluded interaction with the outpatient or inpatient healthcare system.

Finally, our calibration measures may be optimistic due to cross-validation rather than an external validation sample.

CONCLUSION

This study is the first to validate classification algorithms using ICD-10 codes and Medicare administrative data to identify patients with delirium when using large datasets. The algorithms performed well against a real-world reference standard based on expert clinician adjudication of information contained within electronic health records. A validated classification of patients with delirium could improve understanding of delirium at a national or population level and help assess the impact of policies, delivery changes, and clinical interventions to prevent delirium in our older patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Study Funding:

R01AG062282, K08AG053380, R01AG073410, P01AG032952, and U01AG076478

Financial Disclosures:

L.M.V.R.M.: Support from the Centers for Diseases Control and Prevention (U48DP006377), the National Institutes of Health (NIH-NIA 5K08AG053380-02, NIH-NIA 5R01AG062282-02, NIH-NIA 2P01AG032952-11, NIH- NIA 3R01AG062282-03S1), and the Epilepsy Foundation of America and reports no conflict of interest.

S.Z.: Support from the National Institutes of Health (K23NS114201).

N.M.B.: Support from NLM T15 LM007092. Dr. Benson volunteers for the Epic Subspecialty Steering Board.

N.F.: VA Office of Academic Affiliations through the VA/National Clinician Scholars Program and Yale University. Disclaimer: The contents of this manuscript do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Yale National Clinician Scholars Program and by CTSA Grant Number TL1 TR001864 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official view of NIH

M.P.: No conflict of interest to disclose.

M.A.D.: No conflict of interest to disclose.

S.L.N.: Support from the NIH: 1R01AG062282, P01AG032952, and U01AG076478

J.P.N.: Receives funding from NIH (1R01AG062282, P01AG032952, and U01AG076478, T32-AG51108) and reports being a director of Aetna until May 2018 and holding equity in Aetna until November 2018.

D.B.: Support from the NIH (5P30 AG062421-03, 2P01AG036694-11, 5U01AG032984-12, 1U24NS100591-04, 1R01AG058063-04, R01AG063975-03, 5R01AG062282-04, 3R01AG062282-03S1, 5R01AG066793-02, 1U19AG062682-03, 2P01AG032952-11, 2T32MH017119-34 Billing Agreement 010289.0001, 3P01AG032952-12S3, 1U01AG068221-01, 1U01AG076478-01, 5R01AG048351-05) and 747021.Blacker.2019 from President and Fellows of Harvard College reports no conflict of interest.

J.H.: Support from the NIH (1R01AG062282, P01AG032952, and U01AG076478), and reports no conflict of interest

REFERENCES

1. Hshieh TT, Inouye SK, Oh ES. Delirium in the Elderly. Psychiatr Clin North Am. 2018;41(1):1–17. [PubMed: 29412839]
2. Dittrich T, Tschudin-Sutter S, Widmer AF, Ruegg S, Marsch S, Sutter R. Risk factors for new-onset delirium in patients with bloodstream infections: independent and quantitative effect of catheters and drainages-a four-year cohort study. Ann Intensive Care. 2016;6(1):104. [PubMed: 27797086]
3. Lundstrom M, Edlund A, Karlsson S, Brannstrom B, Bucht G, Gustafson Y. A multifactorial intervention program reduces the duration of delirium, length of hospitalization, and mortality in delirious patients. Journal of the American Geriatrics Society. 2005;53(4):622. [PubMed: 15817008]
4. Davis DH, Kreisel SH, Muniz Terrera G, et al. The epidemiology of delirium: challenges and opportunities for population studies. Am J Geriatr Psychiatry. 2013;21(12):1173–1189. [PubMed: 23907068]
5. Blazer DG, van Nieuwenhuizen AO. Evidence for the diagnostic criteria of delirium: an update. Curr Opin Psychiatry. 2012;25(3):239–243. [PubMed: 22449764]
6. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. Lancet. 2014;383(9920):911–922. [PubMed: 23992774]
7. Davis D, Searle SD, Tsui A. The Scottish Intercollegiate Guidelines Network: risk reduction and management of delirium. Age Ageing. 2019;48(4):485–488. [PubMed: 30927352]
8. Goldberg TE, Chen C, Wang Y, et al. Association of Delirium With Long-term Cognitive Decline: A Meta-analysis. JAMA Neurol. 2020;77(11):1373–1381. [PubMed: 32658246]
9. Heriot NR, Levinson MR, Mills AC, Khine TT, Gellie AL, Sritharan G. Diagnosing delirium in very elderly intensive care patients. Intensive Crit Care Nurs. 2017;38:10–17. [PubMed: 27600028]
10. Colby SL, Ortman JM. The Baby Boom Cohort in the United States: 2012 to 2060. United States Census Bureau. https://www.census.gov/library/publications/2014/demo/p25-1141.html. Published 2014. Updated October 8, 2021. Accessed April, 2022.
11. Gou RY, Hshieh TT, Marcantonio ER, et al. One-Year Medicare Costs Associated With Delirium in Older Patients Undergoing Major Elective Surgery. JAMA Surg. 2021;156(5):430–442.
12. Caplan GA, Teodorczuk A, Streatfeild J, Agar MR. The financial and social costs of delirium. Eur Geriatr Med. 2020;11(1):105–112. [PubMed: 32297239]
13. Moura Junior V, Westover MB, Li F, et al. Hospital complications among older adults: Better processes could reduce the risk of delirium. Health Serv Manage Res. 2021:9514848211028707.
14. Kinchin I, Mitchell E, Agar M, Trepel D. The economic cost of delirium: A systematic review and quality assessment. Alzheimers Dement. 2021;17(6):1026–1041. [PubMed: 33480183]
15. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK. One-year health care costs associated with delirium in the elderly population. Arch Intern Med. 2008;168(1):27–32. [PubMed: 18195192]
16. Dharmarajan K, Swami S, Gou RY, Jones RN, Inouye SK. Pathway from Delirium to Death: Potential In-Hospital Mediators of Excess Mortality. J Am Geriatr Soc. 2017;65(5):1026–1033. [PubMed: 28039852]
17. Inouye SK, Bogardus ST Jr., Charpentier PA, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med. 1999;340(9):669–676. [PubMed: 10053175]
18. Inouye SK, Westendorp RGJ, Saczynski JS. Review: Delirium in elderly people. The Lancet. 2014;383:911–922.
19. Israni J, Lesser A, Kent T, Ko K. Delirium as a predictor of mortality in US Medicare beneficiaries discharged from the emergency department: a national claims-level analysis up to 12 months. BMJ Open. 2018;8(5):e021258.
20. Fu S, Lopes GS, Pagali SR, et al. Ascertainment of delirium status using natural language processing from electronic health records. J Gerontol A Biol Sci Med Sci. 2020.
21. Netzer M, Hackl WO, Schaller M, Alber L, Marksteiner J, Ammenwerth E. Evaluating Performance and Interpretability of Machine Learning Methods for Predicting Delirium

in Gerontopsychiatric Patients. Stud Health Technol Inform. 2020;271:121–128. [PubMed: 32578554]

22. Kolanowski AM, Hill NL, Kurum E, et al. Gender differences in factors associated with delirium severity in older adults with dementia. Arch Psychiatr Nurs. 2014;28(3):187–192. [PubMed: 24856271]
23. Moura L, Festa N, Price M, et al. Identifying Medicare beneficiaries with dementia. J Am Geriatr Soc. 2021;69(8):2240–2251. [PubMed: 33901296]
24. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA.: American Psychiatry Association; 2013.
25. Kim DH, Lee J, Kim CA, et al. Evaluation of algorithms to identify delirium in administrative claims and drug utilization database. Pharmacoepidemiol Drug Saf. 2017;26(8):945–953. [PubMed: 28485014]
26. Moura L, Smith JR, Blacker D, et al. Epilepsy Among Elderly Medicare Beneficiaries: A Validated Approach to Identify Prevalent and Incident Epilepsy. Med Care. 2019;57(4):318–324. [PubMed: 30762723]
27. Miller RR 3rd, Ely EW. Delirium and cognitive dysfunction in the intensive care unit. Semin Respir Crit Care Med. 2006;27(3):210–220. [PubMed: 16791755]
28. Rich K Article Review: One-year medicare costs associated with delirium in older patients undergoing major elective surgery by Gou, R., Hshieh, T., Marcantonio, E., et al. JAMA 2021. J Vasc Nurs. 2021;39(2):47–48. [PubMed: 34120698]
29. Vasilevskis EE, Chandrasekhar R, Holtze CH, et al. The Cost of ICU Delirium and Coma in the Intensive Care Unit Patient. Med Care. 2018;56(10):890–897. [PubMed: 30179988]
30. Pezzullo L, Streatfeild J, Hickson J, Teodorczuk A, Agar MR, Caplan GA. Economic impact of delirium in Australia: a cost of illness study. BMJ Open. 2019;9(9):e027514.
31. Schubert M, Schurch R, Boettger S, et al. A hospital-wide evaluation of delirium prevalence and outcomes in acute care patients - a cohort study. BMC Health Serv Res. 2018;18(1):550. [PubMed: 30005646]
32. Levinson DR. Adverse Events in Hospitals: National Incidence Among Medicare Beneficiaries, OEI-06-09-00090. Agency for Healthcare Research and Quality; July 2010 2010.
33. Hospital-Acquired Condition Reduction Program. Centers for Medicare & Medicaid Services. Acute Inpatient PPS Web site. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/HAC-Reduction-Program. Published 2021. Updated 08/09/2021. Accessed 11/23/2021, 2021.
34. Katznelson R, Djaiani G, Tait G, et al. Hospital administrative database underestimates delirium rate after cardiac surgery. Can J Anaesth. 2010;57(10):898–902. [PubMed: 20645040]
35. Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST Jr., Leslie DL, Agostini JV. A chart-based method for identification of delirium: validation compared with interviewer ratings using the confusion assessment method. J Am Geriatr Soc. 2005;53(2):312–318. [PubMed: 15673358]
36. Thomas C, Kreisel SH, Oster P, Driessen M, Arolt V, Inouye SK. Diagnosing delirium in older hospitalized adults with dementia: adapting the confusion assessment method to international classification of diseases, tenth revision, diagnostic criteria. J Am Geriatr Soc. 2012;60(8):1471–1477. [PubMed: 22881707]
37. Swan JT, Fitousis K, Hall JB, Todd SR, Turner KL. Antipsychotic use and diagnosis of delirium in the intensive care unit. Crit Care. 2012;16(3):R84. [PubMed: 22591601]
38. Moura L, Festa N, Price M, et al. Identifying Medicare beneficiaries with dementia. J Am Geriatr Soc. 2021.
39. Sepulveda E, Franco JG, Trzepacz PT, et al. Delirium diagnosis defined by cluster analysis of symptoms versus diagnosis by DSM and ICD criteria: diagnostic accuracy study. BMC Psychiatry. 2016;16:167. [PubMed: 27229307]
40. Bui LN, Pham VP, Shirkey BA, Swan JT. Effect of delirium motoric subtypes on administrative documentation of delirium in the surgical intensive care unit. J Clin Monit Comput. 2017;31(3):631–640. [PubMed: 27108353]

41. Callahan A, Shah NH, Chen JH. Research and Reporting Considerations for Observational Studies Using Electronic Health Record Data. Ann Intern Med. 2020;172(11 Suppl):S79–S84. [PubMed: 32479175]

42. CMS Fast Facts. Centers for Medicare & Medicaid. Research, Statistics, Data & Systems Web site. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Fast-Facts. Published 2021. Updated 08/03/2021. Accessed 11/29, 2021.

43. Gaudette E, Tysinger B, Cassil A, Goldman DP. Health and Health Care of Medicare Beneficiaries in 2030. Forum Health Econ Policy. 2015;18(2):75–96. [PubMed: 27127455]

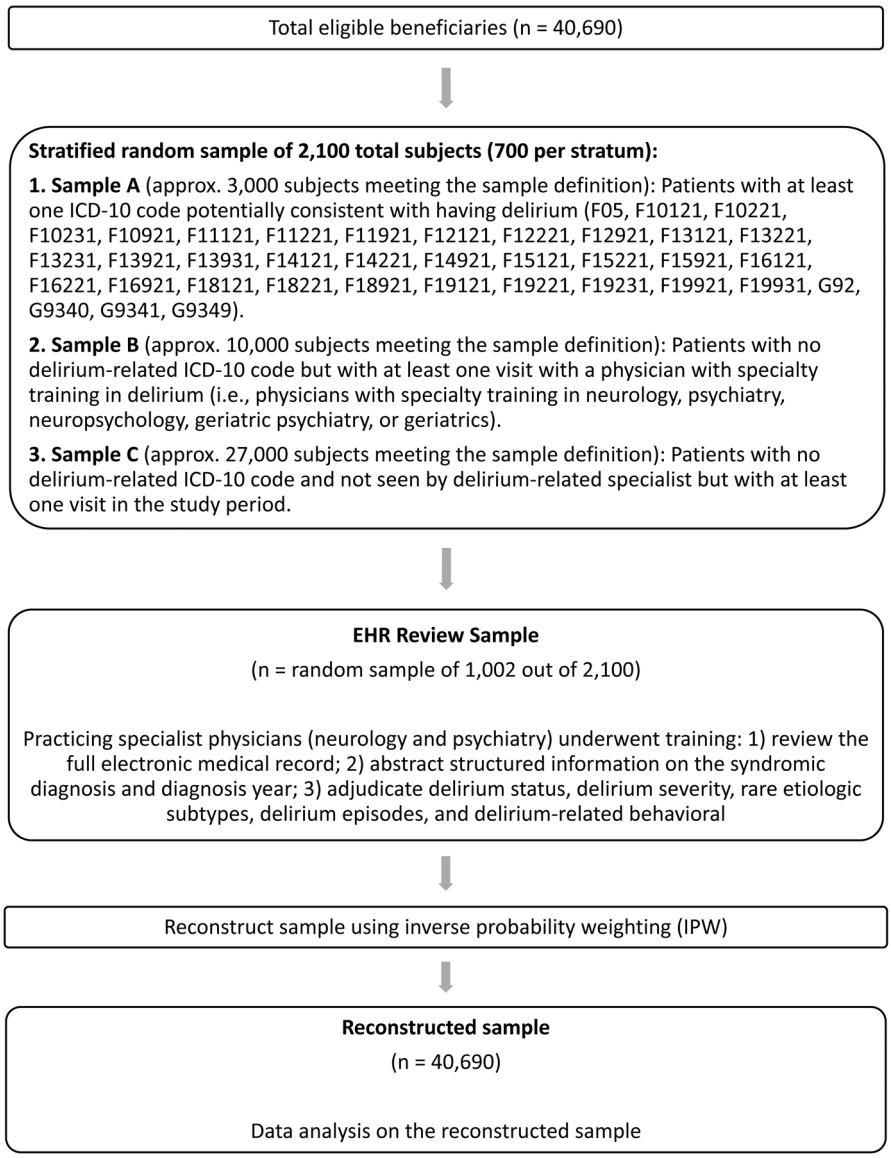

Figure 1. Sampling Approach

The workflow delineates our sampling procedure to build our final reconstructed sample for analysis, beginning with 40,690 FFS Medicare beneficiaries aged 65-years and older within the ACO. We perform stratified random sampling based on the pretest likelihood based on administrative data.

Abbreviations: EHR, electronic health record; ICD-10, International Classification of Diseases, 10th Edition; IPW, inverse proportional weights.

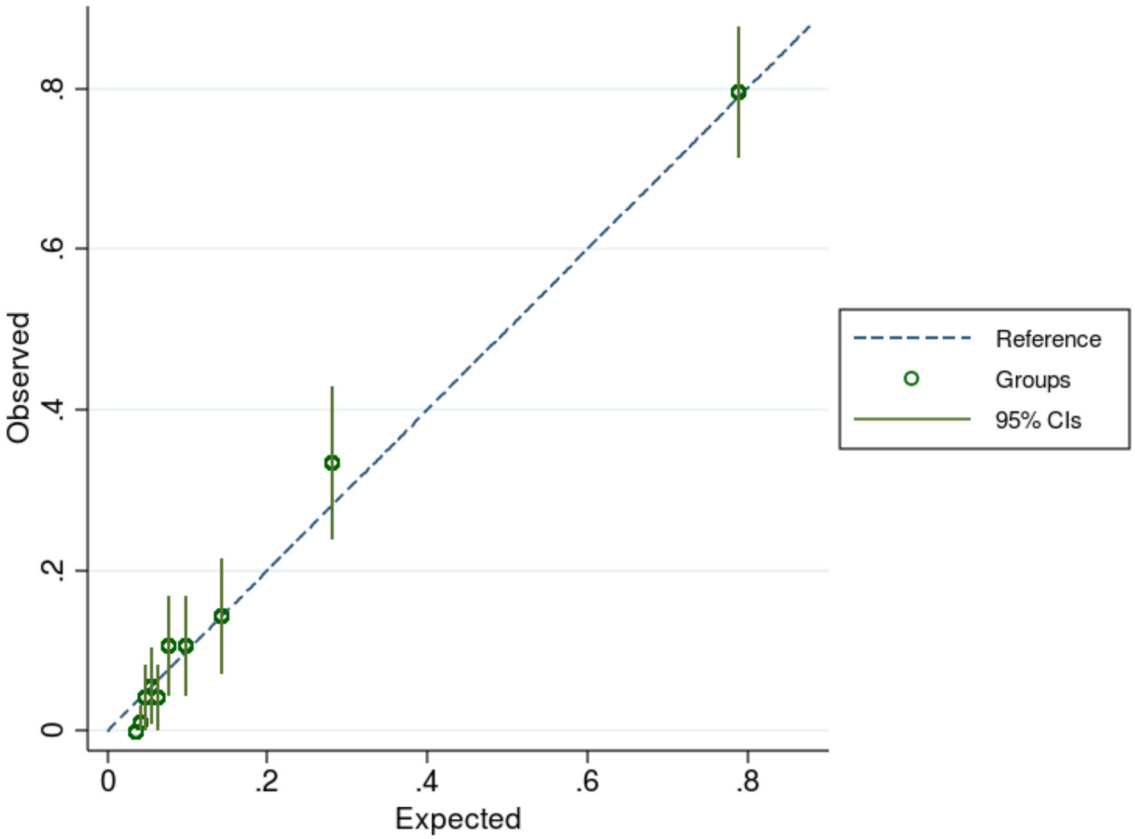

Figure 2. Calibration Plot (Model 8)

We grouped patients by their deciles of estimated probability and plotted the mean of each group's expected (predicted) vs. observed (whether they had delirium). If the model were perfectly calibrated, the cross-validated slope would be 1, the calibration-in-the-large would be 0, and all points would lie on the dashed 45-degree line.

Abbreviations: CV, cross-validation; CITL, calibration-in-the-large

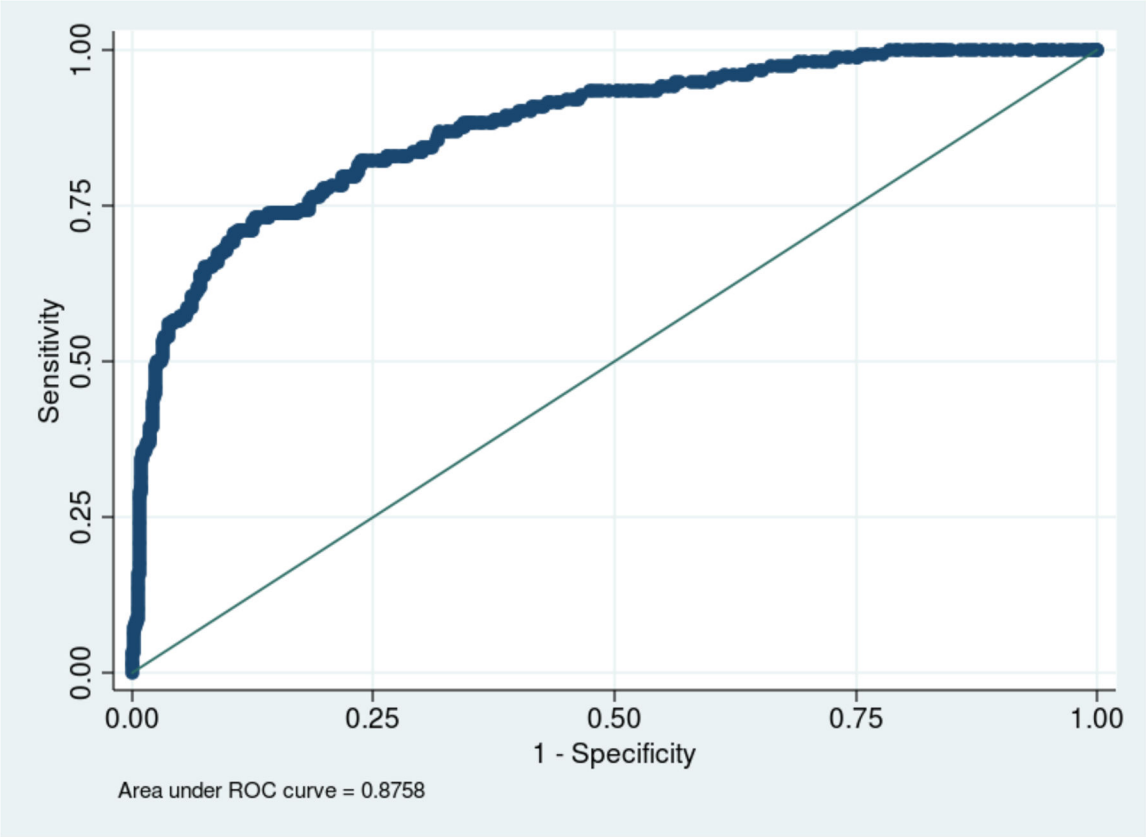

Figure 3. Area Under the ROC curve: Tradeoff Between Sensitivity and Specificity

The figure is based on the analytic sample. Area-under-receiver-operating-curve (AUROC) = 0.87. Figure 3 provides the AUROC related to varying sensitivity and 1-specificity and illustrates the point that AUROC is maximized. We display in Figure 3 the performance reference to Model 8 (i.e., best simple model), which used the "Refined" diagnosis codes, plus a count of hospitalizations (inpatient stays) and outpatient visits with delirium diagnoses, count of delirium-associated (antipsychotic) drug fills, and indicators for dementia status and age.

Abbreviations: AUROC, area-under-receiver-operating-curve

Table 1.

Characteristics of Medicare Beneficiaries Overall, and by Sample

Characteristics	Sample A	Sample B	Sample C	Total
Subjects (n)	2,900	10,657	27,133	40,690
Age (years), mean (SD)	81 (0.41)	76 (0.36)	74 (0.36)	75 (0.260
Female	64%	59%	58%	59%
White	92%	94%	92%	93%
Limited available clinical data in the EHR	7%	3%	5%	5%
Dementia Status	57%	1%	0%	5%
Mean baseline CMS-HCC prospective risk score	1.3 (0.05)	1.2 (0.05)	0.8 (0.04)	0.9 (0.03)
Clinical events during the observation period				
ED visit or hospitalization	84%	77%	48%	58%
Antipsychotic drug receipt	18%	3%	1%	3%
Deaths	7%	3%	2%	2%

Sample A: patients with an ICD-10 diagnosis code suggestive of a cognitive concern; Sample B: patients with a clinic visit during the observation period with a neurologist, psychiatrist, geriatrician, neuropsychologist, or geriatric psychiatrist (i.e., a clinician with specialty training in delirium care), but not in Sample A; Sample C: all other patients. We used a previously validated claims-based algorithm to identify dementia status. The Medicare program uses the Centers for Medicare and Medicaid Service's Hierarchical Condition Category (CMS-HCC) prospective risk score based on year-one diagnoses to predict year-two spending to risk adjust payments under the Medicare Advantage program.

Abbreviations: CMS-HCC, Centers for Medicare and Medicaid Service's Hierarchical Condition Category; ED, Emergency Department; EHR, electronic health record; SD, standard deviation

Table 2.

Characteristics of Beneficiaries without and with EHR-detected Delirium (2016–2018)

	No delirium detected N=35,534	Delirium episode N=2,461
Mean age (SD)	75 (0.28)	81 (0.86)
65–74 years of age	57%	22%
75–79 years of age	21%	21%
80–84 years of age	13%	18%
85+ years of age	9%	39%
Female	58%	63%
White	93%	95%
Mean baseline CMS-HCC risk score (SD)	0.9 (0.03)	1.7 (0.14)
Dementia Status	3%	26%
Clinical events during the observation period		
ED visit or hospitalization	55%	99%
Antipsychotic drug receipt	1%	20%
Death	2%	8%

Expert clinicians adjudicated delirium status after a structured review of all clinical information contained in each patient's EHR. We used a previously validated claims-based algorithm to identify dementia status. The Medicare program uses the Centers for Medicare and Medicaid Service's Hierarchical Condition Category (CMS-HCC) prospective risk score based on year-one diagnoses to predict year-two spending to risk adjust payments under the Medicare Advantage program.

Abbreviations: CMS-HCC, Centers for Medicare and Medicaid Service's Hierarchical Condition Category; ED, Emergency Department; SD, standard deviation