Journal of Clinical and Translational Science

www.cambridge.org/cts

Identification of delirium from real-world electronic health record clinical notes

Jennifer St. Sauver[1,2], Sunyang Fu[3], Sunghwan Sohn[3], Susan Weston[3], Chun Fan[4], Janet Olson[1], Bjoerg Thorsteinsdottir[5], Nathan LeBrasseur[6,7], Sandeep Pagali[5], Walter Rocca[1,8,9] and Hongfang Liu[1,3]

[1]Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; [2]The Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA; [3]Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA; [4]Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; [5]Department of Medicine, Mayo Clinic, Rochester, MN, USA; [6]Robert and Arlene Kogod Center on Aging, Mayo Clinic, Rochester, MN, USA; [7]Department of Physical Medicine and Rehabilitation, Mayo Clinic, Rochester, MN, USA; [8]Department of Neurology, Mayo Clinic, Rochester, MN, USA and [9]Women's Health Research Center, Mayo Clinic, Rochester, MN, USA

Translational Research, Design and Analysis Research Article

Cite this article: St. Sauver J, Fu S, Sohn S, Weston S, Fan C, Olson J, Thorsteinsdottir B, LeBrasseur N, Pagali S, Rocca W, and Liu H. Identification of delirium from real-world electronic health record clinical notes. *Journal of Clinical and Translational Science* **7**: e187, 1–7. doi: 10.1017/cts.2023.610

Received: 25 January 2023
Revised: 2 August 2023
Accepted: 8 August 2023

Keywords:
Delirium; natural language processing algorithm; International Classification of Diseases (ICD); electronic health records; bioinformatics

Corresponding author:
J. L. St. Sauver, PhD;
Email: stsauver.jennifer@mayo.edu

Abstract

Introduction: We tested the ability of our natural language processing (NLP) algorithm to identify delirium episodes in a large-scale study using real-world clinical notes. *Methods:* We used the Rochester Epidemiology Project to identify persons ≥ 65 years who were hospitalized between 2011 and 2017. We identified all persons with an International Classification of Diseases code for delirium within ±14 days of a hospitalization. We independently applied our NLP algorithm to all clinical notes for this same population. We calculated rates using number of delirium episodes as the numerator and number of hospitalizations as the denominator. Rates were estimated overall, by demographic characteristics, and by year of episode, and differences were tested using Poisson regression. *Results:* In total, 14,255 persons had 37,554 hospitalizations between 2011 and 2017. The code-based delirium rate was 3.02 per 100 hospitalizations (95% CI: 2.85, 3.20). The NLP-based rate was 7.36 per 100 (95% CI: 7.09, 7.64). Rates increased with age (both $p < 0.0001$). Code-based rates were higher in men compared to women ($p = 0.03$), but NLP-based rates were similar by sex ($p = 0.89$). Code-based rates were similar by race and ethnicity, but NLP-based rates were higher in the White population compared to the Black and Asian populations ($p = 0.001$). Both types of rates increased significantly over time (both p values < 0.001). *Conclusions:* The NLP algorithm identified more delirium episodes compared to the ICD code method. However, NLP may still underestimate delirium cases because of limitations in real-world clinical notes, including incomplete documentation, practice changes over time, and missing clinical notes in some time periods.

Introduction

Delirium is a common, disorienting condition in hospitalized patients and represents a significant management challenge for health care staff [1,2]. Delirium is also associated with prolonged length of stay, an increased risk of institutional discharge and 30-day readmission, long-term cognitive decline, and mortality [3–8]. Thus, the overall monetary and societal costs related to delirium are substantial [9].

A recent meta-analysis suggests that the incidence of delirium in hospitalized adult patients remained relatively stable between 1980 and 2019, with a pooled cumulative incidence of 9% [10]. However, delirium is routinely underdiagnosed, particularly mild cases, and electronic billing codes incompletely capture this condition [11–14]. Underdiagnosis with billing codes represents a significant barrier to conducting retrospective studies to understand the natural history of delirium and to determine if delirium incidence is changing over time. In addition, incomplete identification of delirium can substantially hamper clinical research efforts that use large databases to identify risk factors for and outcomes of delirium [11].

Although billing codes in administrative datasets may incompletely identify delirium cases, clinical notes frequently contain details that are relevant to a delirium diagnosis. Therefore, we have previously developed a natural language processing (NLP) algorithm to identify delirium episodes from electronic health record (EHR) clinical notes [15,16] based on the confusion assessment method (CAM) framework [17,18]. The NLP delirium algorithm had high sensitivity (92%) and specificity (100%) for identification of delirium from the clinical notes of persons participating in the Mayo Clinic Biobank [15], and captured 80% of delirium cases in patients hospitalized with COVID-19 [19]. However, the algorithm has not yet been tested in a large-scale study using real-world clinical notes derived from more than one health care

institution. The goal of this study was to assess the ability of the NLP algorithm to identify episodes of delirium in a large, general population using notes from multiple health care institutions. Therefore, we applied this algorithm to the medical records of hospitalized adults (≥ 65 years) residing in Olmsted County, Minnesota over a seven-year time period. We compared the cases identified via the algorithm to cases identified using International Classification of Diseases (ICD) billing codes. We also examined temporal trends in rates of delirium over time and by age, sex, race, and ethnicity using these methods.

Materials and Methods

Study population and data source. We used the resources of the Rochester Epidemiology Project (REP) records-linkage system to identify persons 65 years and older residing in Olmsted County, MN who were hospitalized at least once between 2011 and 2017. Persons with at least one hospitalization during this time frame were considered "at risk" for a delirium episode.

The REP has been previously described [20,21]. Briefly, the REP includes linked medical records from local health care institutions (Olmsted Medical Center and satellite clinics, Mayo Clinic and satellite clinics, Olmsted County Public Health Services, and Zumbro Valley Health) for persons who have lived in Olmsted County, MN since 1966. Through this collaboration across health care institutions, the REP captures virtually all of the health care information delivered to the population residing in Olmsted County, MN [21]. Health care data from all visits to each health care institution are coded and indexed electronically, and the full text of the clinical notes from each provider is available for NLP studies. This study was approved by the Mayo Clinic (#18-006044) and Olmsted Medical Center Institutional Review Boards (#035-OMC-18).

Definitions of Delirium

International Classification of Diseases code-based. Among persons 65 years old or older with a hospitalization during the study period, we identified all persons with an ICD code for delirium within ±14 days of a hospital admission (Supplemental Table 1). Forty-five persons had hospitalizations 14–27 days apart, and we reviewed the records for a random sample of 10 persons. All had a single hospitalization episode; therefore, we classified all hospitalizations ≤ 28 days apart as a single episode. Codes from hospitalizations separated by >28 days were considered separate episodes. The date when the first code was assigned within an episode was used as the episode date.

Natural language processing-algorithm-based. Our rule-based NLP algorithm was previously developed to automatically review clinical notes to identify patients with delirium based on the confusion assessment method (CAM) criteria [15,16]. Briefly, we developed specific review criteria ("annotation guidelines;" Supplemental Materials) to serve as the gold standard for defining delirium episodes. Trained annotators reviewed and annotated EHRs to determine whether the clinical notes contained information that would satisfy the CAM criteria for a delirium episode. The NLP delirium algorithm was then developed to search for delirium-related concepts relevant to the CAM criteria from clinical notes for a given patient and to aggregate the concepts to ascertain a patient's delirium status. The original NLP algorithm developed on Mayo Clinic EHRs was also refined and validated on the EHRs from a second medical center (Olmsted Medical Center; OMC). The original NLP algorithm had a sensitivity of 92% and a specificity of 100% for identifying delirium from Mayo Clinic Rochester notes [15]. Before deploying the algorithm to OMC notes, 400 patients were randomly stratified by the presence of delirium ICD-9 code. Clinical notes from the cohort were manually reviewed ("annotated") to identify whether the notes contained information that met the CAM criteria for delirium assessment. These manually reviewed notes served as the gold standard for identification of delirium. We have previously reported the details of the CAM criteria [15] and the Supplemental Appendix contains details describing the definition of the CAM-related clinical concepts and the specific review guidelines for identification of delirium through manual review of clinical notes ("Annotation guideline"). We then used the first half of the OMC data to refine the NLP algorithm and the second half as independent blind test data. After refinement, the NLP algorithm achieved sensitivity and specificity of 100% and 100% on the OMC test data. Additional implementation details, evaluation results, and source code can be found in Supplemental Appendix.

Clinical notes from the two primary health care institutions (Mayo Clinic and OMC) for the study population were extracted and the refined algorithm was applied to all notes between 2011 and 2017. This time frame was chosen because the NLP algorithm was developed on EHR records prior to an EHR system conversion in 2017. Specifically, Mayo Clinic converted from the GE Centricity EHR to Epic, and OMC converted from the Cerner system to Epic in 2018. The NLP algorithm has not been specifically tested in the Epic clinical notes [15]. An anomaly in back-loads of clinical data in 2016 resulted in the complete availability of clinical notes only from July 1 through December 31, 2016 (6 months). To adjust for this anomaly, all cases of delirium identified during this time period were counted twice to estimate the total number of cases for 2016.

Agreement between the two methods. We studied the ability of the two methods to identify delirium episodes by manually reviewing a stratified random sample of EHRs from 200 persons and separately assessing the accuracy of the ICD code method and the NLP algorithm method. Specifically, we reviewed a random sample of records for 50 persons identified as having delirium by both methods, 50 persons identified as having delirium by ICD code, but not NLP, 50 persons identified as having delirium by NLP but not ICD code, and 50 persons not identified as having delirium by either method. Manual review was conducted blinded to the identification method and was performed using the CAM criteria for delirium ascertainment, as previously described [15]. The CAM review criteria ("annotation guideline") are available in the Supplemental Appendix. The stratified sample selected for manual review was oversampled for delirium cases, but this approach does not impact sensitivity and specificity [22].

Analysis. Rates of delirium were calculated using number of delirium episodes as the numerator and number of hospitalizations as the denominator. Rates were determined for code-based episodes and NLP-based episodes. Rates were calculated overall and stratified by age group (65–69, 70–74, 75–79, 80–84, ≥ 85), sex, race (White, Black, Asian, other/mixed), ethnicity (Hispanic, non-Hispanic), and calendar year of delirium episode. Poisson regression was used to test for differences in delirium rates by each of the patient characteristics. The number of delirium episodes was used as the dependent variable and the natural logarithm of the number of hospitalizations as the independent variable. To test for a trend over time in the rate of delirium, calendar year was modeled as a continuous variable, and the

https://doi.org/10.1017/cts.2023.610 Published online by Cambridge University Press

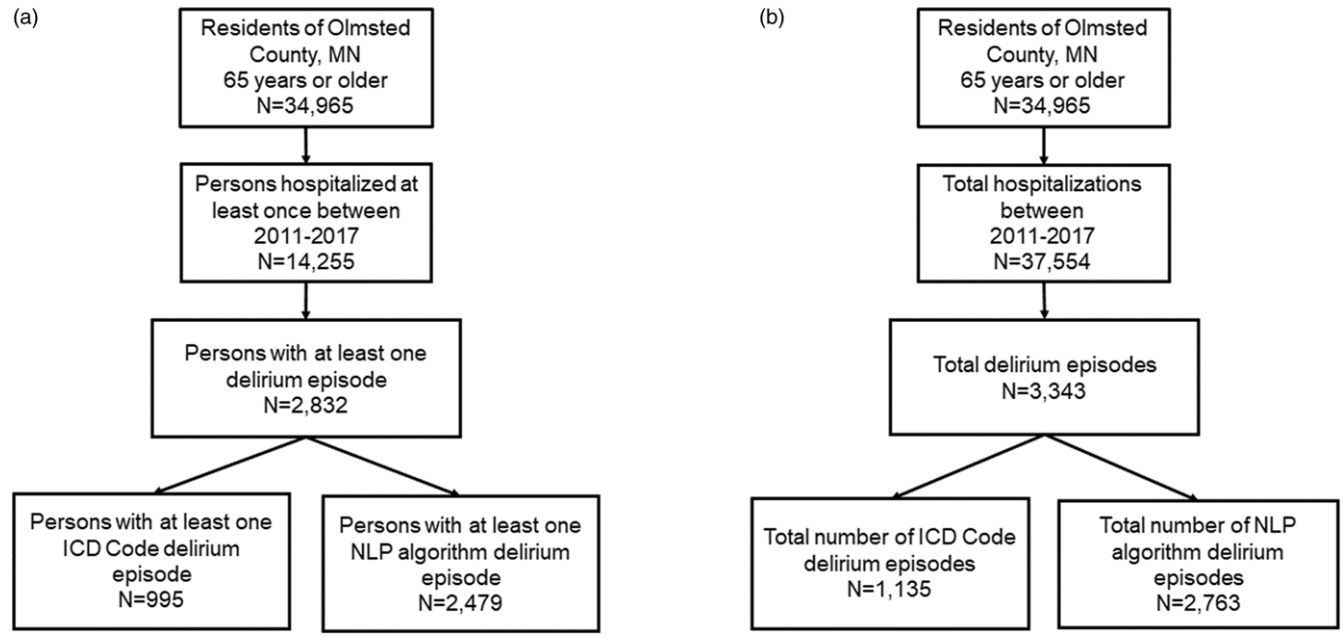

Figure 1. Study population, hospitalizations, and delirium events. ***a***) Indicates the total number of persons with at least one delirium episode, and ***b***) indicates the total number of delirium episodes that occurred during a hospitalization.

models were adjusted for age and sex. Linear and nonlinear trends were tested using Poisson regression by including year and year [2] in the model. P values<0.05 were considered statistically significant.

We also examined agreement between the code and NLP methods for identifying delirium. We calculated overall agreement and a kappa statistic to assess concurrence between the methods. Sensitivity and specificity of each method were calculated by comparing the ascertainment method (ICD code or NLP algorithm) results to manual chart review of the medical records for a stratified sample of 200 persons.

Results

Overall, 34,965 persons 65 years of age or older resided in Olmsted County, MN, and 14,255 had at least one hospitalization between 2011 and 2017 (Fig. 1a). Persons could have multiple hospitalizations, and there were a total of 37,554 hospitalizations during the time frame (approximately 2.6 hospitalizations per patient; Fig. 1b). Among persons who were hospitalized, 2,832 persons had 3,343 delirium episodes s identified either by ICD code or by NLP algorithm (Fig. 1). Overall, 995 persons had at least one ICD code-based delirium episode (1,135 total episodes), and 2,479 persons had at least one NLP-based episode between 2011 and 2017 (2,763 total episodes; Fig. 1).

Characteristics of persons identified with a code-based delirium episode and those with an NLP-based delirium episode are shown in Table 1. The NLP algorithm consistently identified more cases of delirium than the code-based method in all age, sex, racial, and ethnic groups. Table 1 also indicates the rate of delirium episodes identified by each method. Overall, the code-based delirium episode rate was 3.02 per 100 hospitalizations (95% CI: 2.85, 3.20). The NLP-based episode rate was over 2 times higher (7.36 per 100; 95% CI: 7.09, 7.64). Rates of delirium episodes identified by either method increased with age (both $p < 0.001$). Code-based rates were higher in men compared to women ($p = 0.03$), but NLP-based rates were similar by sex ($p = 0.89$; Table 1). Code-based delirium rates were similar by race and ethnicity, but NLP-based rates were significantly higher in the White population compared to the Black and Asian populations (Table 1). Code-based delirium rates increased significantly over time, with a particular increase between 2015 and 2016 (p value for year [2]<0.001, adjusted for age and sex; Table 1 and Fig. 2). Similarly, NLP-based delirium rates also increased significantly over time (p value for year[2] <0.001, adjusted for age and sex; Table 1 and Fig. 2).

Finally, we assessed the agreement, sensitivity, and specificity of the two methods for identifying delirium. Overall agreement between the ICD code and NLP methods for identifying delirium was 86%, but kappa was 34% (95% CI: 31%, 36%; indicating fair agreement). Compared to manual review of a stratified sample of medical records for 200 persons, the ICD code method for delirium identification had a sensitivity = 60% (95% CI: 52%, 68%) and a specificity = 73% (95% CI: 61%, 85%). The NLP method had a sensitivity of 64% (95% CI: 56%, 72%) and a specificity of 84% (95% CI: 74%, 93%). While the sensitivity of the two methods was similar, overlap between the two methods was modest. Overall, 45 true cases (compared to manual review) were identified by both methods. The ICD code method identified 39 additional true cases that were not identified by the NLP algorithm. By contrast, the NLP method identified 45 additional true cases that were not identified by the ICD code method (Supplemental Figure 1).

Discussion

We used an ICD-code-based method and an NLP algorithm with high sensitivity and specificity to identify delirium episodes from the real-world clinical notes of a population-based sample of hospitalized patients between 2011 and 2017. We found an increase in delirium rates over time using both methods; however, the NLP algorithm consistently identified more delirium episodes compared to the ICD code method. We also identified several important problems to consider when applying our NLP algorithm

https://doi.org/10.1017/cts.2023.610 Published online by Cambridge University Press

Table 1. Number of delirium episodes and rates of delirium cases detected using International Classification of Diseases (ICD) codes or a natural language processing (NLP) algorithm

		ICD codes			NLP algorithm		
	Hospitalizations *N*	Delirium episodes *N*	Rate per 100 hospitalizations (95% CI[+])	*P* value	Delirium episodes* *N*	Rate per 100 hospitalizations (95% CI[+])	*P* value
Overall	37,554	1,135	3.02 (2.85,3.20)		2,763	7.36 (7.09,7.64)	
Age group				<0.0001			<0.0001
65-69	6,759	116	1.72 (1.42,2.06)		248	3.67 (3.23,4.16)	
70-74	6,850	155	2.26 (1.92,2.65)		294	4.29 (3.82,4.81)	
75-79	6,973	170	2.44 (2.09,2.83)		412	5.91 (5.35,6.51)	
80-84	6,707	247	3.68 (3.24,4.17)		556	8.29 (7.62,9.01)	
≥85	10,265	447	4.35 (3.96,4.78)		1,253	12.21 (11.54,12.90)	
Sex				0.03			0.89
Men	17,945	578	3.22 (2.96,3.49)		1,324	7.38 (6.99,7.79)	
Women	19,609	557	2.84 (2.61,3.09)		1,439	7.34 (6.96,7.73)	
Race				0.19			0.001
White	34,111	1,040	2.66 (1.52,4.32)		2,569	7.53 (7.24,7.83)	
Black	602	16	2.36 (1.72,3.16)		30	4.98 (3.36,7.11)	
Asian	1,905	45	3.74 (2.59,5.23)		107	5.62 (4.60,6.79)	
Other/mixed	909	34	3.05 (2.87,3.24)		57	6.27 (4.75,8.12)	
Hispanic ethnicity				0.82			0.20
No	36,790	1,113	3.03 (2.85,3.21)		2,716	7.38 (7.11,7.67)	
Yes	764	22	2.88 (1.80,4.36)		47	6.15 (4.52,8.18)	
Year				<0.0001			<0.0001
2011	4,941	105	2.13 (1.74,2.57)		295	5.97 (5.31,6.69)	
2012	5,365	121	2.26 (1.87,2.69)		378	7.05 (6.35,7.79)	
2013	5,104	143	2.80 (2.36,3.30)		366	7.17 (6.45,7.94)	
2014	5,110	121	2.37 (1.96,2.83)		390	7.63 (6.89,8.43)	
2015	5,342	129	2.41 (2.02,2.87)		357	6.68 (6.01,7.41)	
2016	5,796	208	3.59 (3.12,4.11)		484	8.35 (7.34, 9.58)**	
2017	5,896	308	5.22 (4.66,5.84)		493	8.36 (7.64,9.13)	

[+]CI = confidence interval.
*Numbers reflect cloned cases for 2016.
**CI based on using only cases from 7/1/2016-12/31/2016.

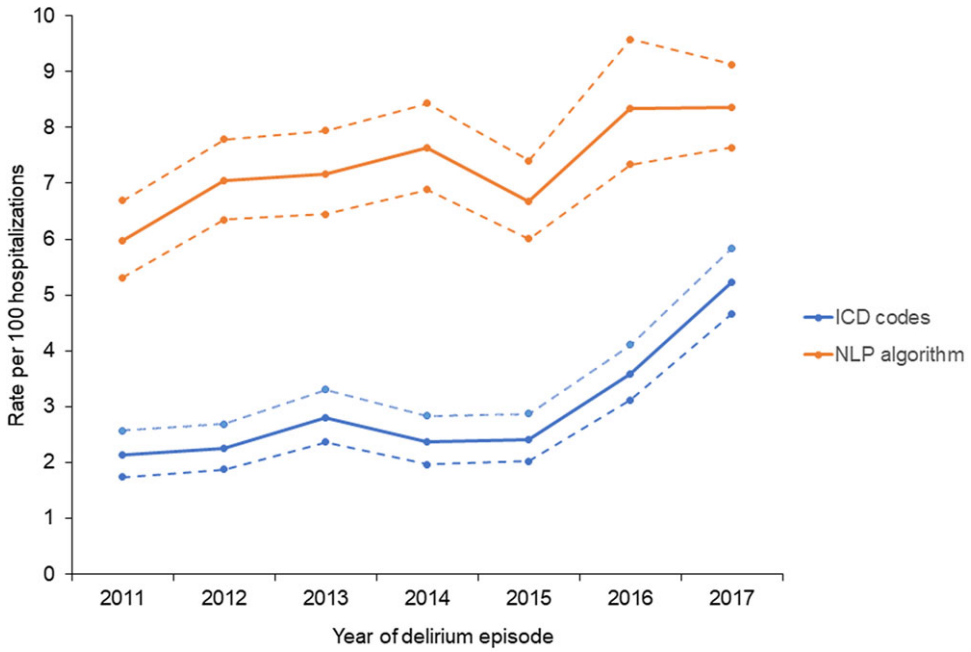

Figure 2. Rates of delirium between 2011 and 2017 using two methods of identification. Rates identified using international classification of diseases (ICD) codes or the natural language processing (NLP) algorithm are displayed with 95% confidence intervals. Rates of detection increased over time using both methods (both *P* value tests for trend < 0.001).

to real-world clinical notes, including incomplete documentation, the impact of practice changes over time, and missing clinical notes in some time periods.

We found that the NLP algorithm consistently identified over twice as many delirium episodes as the ICD codes when applied to clinical notes that were collected as a routine part of clinical care. Identification of delirium using ICD codes has been previously reported as problematic [11–14], and it is not surprising that the NLP algorithm identified more delirium episodes in this population. However, the methods had similar sensitivities compared to manual review (ICD method: 60%; NLP method: 64%), indicating that both methods missed true cases compared to manual review. Therefore, to completely identify delirium cases, a combination of the two methods, followed by medical record review, may provide optimal identification of cases. In addition, algorithm performance relies on documentation in clinical notes to accurately identify cases. If procedures are not in place in the clinical setting to accurately capture delirium episodes, the NLP algorithm will not be able to identify these episodes. We also note that the NLP algorithm sensitivity was lower in this real-world application compared to the 83% sensitivity observed in the initial algorithm development [23]. Such declines in sensitivity are not uncommon in application of algorithms to new samples, but further analysis is needed to ensure optimal algorithm performance in new settings.

We also found significant increases over time in identification of delirium rates using both the code and the NLP methods for case identification. We do not know of a biological reason that would explain an increase in delirium rates during this time period. Additionally, these findings are in contrast to a previous study which found no significant temporal changes in delirium prevalence between 1980 and 2020 [10]. However, our findings are consistent with a study that found an increase in ICD-based diagnoses of delirium and encephalopathy between 2011 and 2018 [24]. These authors noted a significant increase in delirium diagnoses between 2014 and 2016, which they attribute to the national shift from the ICD-9 to the ICD-10 coding system. ICD-10 significantly expanded the number of codes that could be used to identify delirium (Supplemental Table 1). The increase in the available number of codes, combined with an increased recognition of delirium as an important clinical condition in aging populations could account for the increase we observed in code-based delirium diagnoses after 2015. The increase in delirium cases over time identified by the NLP method may also indicate an increased recognition by health care providers that delirium is a significant concern in aging populations. Such recognition may lead to an increased documentation of delirium concepts in clinical notes, and capture of true delirium episodes may have improved in more recent years. Unfortunately, we do not have detailed information on changes to clinical practice that could impact changes to delirium documentation during this time frame. Our results highlight an important issue encountered when applying NLP algorithms to real-world data. NLP algorithm performance is dependent on clinical note documentation, and documentation may change over time with changes in clinical practice. Therefore, when using NLP algorithms, it may be important to limit study time frames to the most recent clinical notes to ensure that practice and documentation changes do not affect study results. In addition, it is important for investigators to understand that using real-world EHR data requires considering both biologic and non-biologic reasons for changes in disease incidence and prevalence over time.

Our overall NLP-based delirium rate of 7.4% is lower than the pooled prevalence rate of 15% estimated in a recent review and meta-analysis of 33 studies [10]. An additional meta-analysis of 9 studies suggested a prevalence ranging from 9% to 32% [25]. However, our study differs in several ways from many of the studies included in these analyses [10,26]. First, most previous studies directly examined patients to assess delirium after obtaining informed consent. By contrast, delirium was not

https://doi.org/10.1017/cts.2023.610 Published online by Cambridge University Press

routinely assessed in the hospitals that participated in the REP during our study time frame. Therefore, we expect that some cases were not documented in the medical records and would not be captured by either billing codes or through the NLP algorithm. We, therefore, expect our rates to likely underestimate the overall delirium rate in the hospitalized population aged 65 years or older. However, identification of persons with delirium using EHRs is still useful for retrospective clinical research studies as long as the episodes that are identified are representative of all delirium episodes. In particular, most of the studies included in the meta-analyses were not population-based (admission to single, specialized hospitals or wards), and many assessed delirium in severely ill patients (e.g. patients presenting with a stroke diagnosis [27] or with a cancer diagnosis [28,29]). In addition, most of the previous community-based studies with active assessment of delirium [30–34]. have included a high proportion of persons with dementia or cognitive disorders (40%–60% of admitted patients). Overall, 18% of persons with code-based delirium and 15% of persons with NLP-based delirium had a diagnosis of dementia in the 5 years prior to the first delirium episode during the study period. Dementia patients are at higher risk for experiencing a delirium episode; therefore, we would expect our delirium rates to be lower than in these previous studies because we studied a general population with lower rates of dementia. Although we likely missed some delirium cases using diagnostic codes or the NLP algorithm, the delirium patients that were identified with these methods may be more reflective of delirium cases occurring in the general population. As such, results of retrospective case-control or cohort studies that include delirium patients identified from the EHR using ICD code or NLP algorithm methods are likely to be unbiased.

We also observed a significantly higher rate of delirium in our White population compared to our Black and Asian populations when using the NLP algorithm for case identification. We do not expect delirium rates to differ by race or ethnicity in persons of similar ages. However, sociodemographic disparities are likely reflected in medical record documentation [35]. For example, Sun and colleagues found more negative descriptors in the medical record notes of black patients compared to white patients (including "refused," "not adherent," "not compliant," and "agitated") [36]. Disparities may also manifest in less complete documentation in persons with different sociodemographic characteristics. NLP algorithms can only identify cases when documentation is present. If persons of different racial or ethnic groups have different levels of documentation or if the language used for such patients differs from that required by the algorithm, true cases will be missed in these patients. Our findings may therefore be biased because our results are based on clinical notes that reflect bias in real-world clinical practice. Further studies are necessary to test the ability of our NLP algorithm to identify delirium from the EHRs of other medical centers that care for diverse patients.

Strengths of our study include application of a previously developed NLP algorithm with high sensitivity and specificity to real-world clinical notes for a large, general study population served by several health care institutions. In addition, availability of both clinical notes and ICD billing codes offered the opportunity to compare the two methods for identification of delirium in the same population. Limitations of our study include the lack of complete availability of clinical notes for half of 2016 (6 months). If delirium rates were substantially different in the half of the year for which we lacked complete data, our estimate of delirium rates in 2016 would be incorrect. However, we note that although the confidence intervals around our rate for 2016 are wider than if we had complete data in that year, adjusting our rates for the missing data did not suggest a significant over- or under-estimate of rates in that year. In addition, as shown in Table 1 and Fig. 2, the 2016 delirium rates are consistent with a gradual increase in cases over time from 2011 to 2017, and the 2016 rate is virtually identical to the rate observed in 2017. The increase in 2016 cases also matches the trend seen in code-based cases. These results suggest that double-counting cases of delirium in 2016 likely resulted in a reasonable estimate of delirium cases in that year. However, this limitation also points to another problem encountered when working with real-world data. Real-world data are routinely affected by technical issues as well as changes in health care policies and procedures. We expect that our study results were affected by the technical issue of missing data in 2016, by changes from the ICD-9 to the ICD-10 coding system, and by changes in documentation of delirium episodes over time. Similarly, both Mayo Clinic and OMC changed their EHR systems to Epic in 2018. The NLP algorithm was developed and tested in the GE Centricity and Cerner clinical notes. We expect performance of the algorithm to be similar in Epic notes, but these studies have not yet been conducted. It is important to recognize these changes and limitations in the interpretation of study results.

In summary, we found that an NLP-based algorithm to identify delirium episodes in a general population using real-world clinical notes may incompletely identify all delirium episodes. The NLP-based algorithm identified more cases compared to ICD codes, and the characteristics of persons identified with delirium are representative of the underlying population most at risk for delirium. Therefore, the use of this algorithm may be appropriate for studies of risk factors and outcomes of delirium.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/cts.2023.610.

Acknowledgments. We thank Trisha Shulze for her assistance with manuscript preparation and submission.

Funding statement. This study used the resources of the REP medical records-linkage system, which is supported by the National Institute on Aging (R33 AG 058738), by the Mayo Clinic Research Committee, and by fees paid annually by REP users. Additional support was provided by R01 AG 72799. The content of this article is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health or the Mayo Clinic.

Competing interests. None.

References

1. **Young J, Inouye SK.** Delirium in older people. *BMJ*. 2007;**334**(7598): 842–846.
2. **Williamson R, Lauricella K, Browning A**, ***et al.*** Patient factors associated with incidents of aggression in a general inpatient setting. *J Clin Nurs*. 2014;**23**(7-8):1144–1152.
3. **Gleason LJ, Schmitt EM, Kosar CM**, ***et al.*** Effect of delirium and other major complications on outcomes after elective surgery in older adults. *JAMA Surg*. 2015;**150**(12):1134–1140.
4. **Leslie DL, Zhang Y, Holford TR, Bogardus ST, Leo-Summers LS, Inouye SK.** Premature death associated with delirium at 1-year follow-up. *Arch Intern Med*. 2005;**165**(14):1657–1662.
5. **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.
6. **Witlox J, Eurelings LS, de Jonghe JF, Kalisvaart KJ, Eikelenboom P, van Gool WA.** Delirium in elderly patients and the risk of postdischarge

mortality, institutionalization, and dementia: a meta-analysis. *JAMA*. 2010;**304**(4):443–451.
7. **Cole MG, Primeau FJ.** Prognosis of delirium in elderly hospital patients. *CMAJ*. 1993;**149**(1):41–46.
8. **Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P.** Does delirium contribute to poor hospital outcomes? A three-site epidemiologic study. *J Gen Intern Med*. 1998;**13**(4):234–242.
9. **Leslie DL, Inouye SK.** The importance of delirium: economic and societal costs. *J Am Geriatr Soc*. 2011;**59**(Suppl 2):S241–3.
10. **Gibb K, Seeley A, Quinn T**, ***et al.*** The consistent burden in published estimates of delirium occurrence in medical inpatients over four decades: a systematic review and meta-analysis study. *Age Ageing*. 2020;**49**(3):352–360.
11. **McCoy TH Jr., Snapper L, Stern TA, Perlis RH.** Underreporting of delirium in statewide claims data: implications for clinical care and predictive modeling. *Psychosomatics*. 2016;**57**(5):480–488.
12. **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.
13. **Elie M, Rousseau F, Cole M, Primeau F, McCusker J, Bellavance F.** Prevalence and detection of delirium in elderly emergency department patients. *CMAJ*. 2000;**163**(8):977–981.
14. **Ritter SRF, Cardoso AF, Lins MMP, Zoccoli TLV, Freitas MPD, Camargos EF.** Underdiagnosis of delirium in the elderly in acute care hospital settings: lessons not learned. *Psychogeriatrics*. 2018;**18**(4):268–275.
15. **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;**77**(3)524–30.
16. **Fu S, Wen A, Pagali S**, ***et al.*** The implication of latent information quality to the reproducibility of secondary use of electronic health records. *Stud Health Technol Inform*. 2022;**290**:173–177.
17. **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.
18. **Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI.** Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med*. 1990;**113**(12):941–948.
19. **Pagali SR, Kumar R, Fu S, Sohn S, Yousufuddin M.** Natural language processing CAM algorithm improves delirium detection compared with conventional methods. *Am J Med Qual*. 2023;**38**(1):17–22.
20. **St Sauver JL, Grossardt BR, Yawn BP, Melton LJ3rd, Pankratz JJ, Brue SM.** Data resource profile: the rochester epidemiology project (REP) medical records-linkage system. *Int J Epidemiol*. 2012;**41**(6):1614–1624.
21. **St Sauver JL, Grossardt BR, Yawn BP, Melton LJ3rd, Rocca WA.** Use of a medical records linkage system to enumerate a dynamic population over time: the rochester epidemiology project. *Am J Epidemiol*. 2011;**173**(9):1059–1068.
22. **Monaghan TF, Rahman SN, Agudelo CW**, ***et al.*** Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. *Medicina (Kaunas)*. 2021;**57**(5):503.
23. **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*. 2022;**77**(3):524–530.
24. **Franks JA, Anderson JL, Bowman E, Li CY, Kennedy RE, Yun H.** Inpatient diagnosis of delirium and encephalopathy: coding trends in 2011-2018. *J Acad Consult Liaison Psychiatry*. 2022;**63**(5):413–422.
25. **Koirala B, Hansen BR, Hosie A**, ***et al.*** Delirium point prevalence studies in inpatient settings: a systematic review and meta-analysis. *J Clin Nurs*. 2020;**29**(13-14):2083–2092.
26. **Siddiqi N, House AO, Holmes JD.** Occurrence and outcome of delirium in medical in-patients: a systematic literature review. *Age Ageing*. 2006;**35**(4):350–364.
27. **Sheng AZ, Shen Q, Cordato D, Zhang YY, Yin Chan DK.** Delirium within three days of stroke in a cohort of elderly patients. *J Am Geriatr Soc*. 2006;**54**(8):1192–1198.
28. **Uchida M, Okuyama T, Ito Y**, ***et al.*** Prevalence, course and factors associated with delirium in elderly patients with advanced cancer: a longitudinal observational study. *Jpn J Clin Oncol*. 2015;**45**(10):934–940.
29. **Grandahl MG, Nielsen SE, Koerner EA, Schultz HH, Arnfred SM.** Prevalence of delirium among patients at a cancer ward: clinical risk factors and prediction by bedside cognitive tests. *Nord J Psychiatry*. 2016;**70**(6):413–417.
30. **Casey P, Cross W, Webb-St Mart M, Baldwin C, Riddell K, Darzins P.** Hospital discharge data under-reports delirium occurrence: results from a point prevalence survey of delirium in a major Australian health service. *Intern Med J*. 2019;**49**(3):338–344.
31. **Bellelli G, Morandi A, Di Santo SG**, ***et al.*** Delirium day": a nationwide point prevalence study of delirium in older hospitalized patients using an easy standardized diagnostic tool. *BMC Med*. 2016;**14**(1):106.
32. **Praditsuwan R, Limmathuroskul D, Assanasen J**, ***et al.*** Prevalence and incidence of delirium in Thai older patients: a study at general medical wards in Siriraj hospital. *J Med Assoc Thai*. 2012;**95** Suppl 2:S245–50.
33. **Yam KK, Shea YF, Chan TC**, ***et al.*** Prevalence and risk factors of delirium and subsyndromal delirium in Chinese older adults. *Geriatr Gerontol Int*. 2018;**18**(12):1625–1628.
34. **Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS.** Impact of different diagnostic criteria on prognosis of delirium: a prospective study. *Dement Geriatr Cogn Disord*. 2004;**18**(3-4):240–244.
35. **London AJ.** Artificial intelligence in medicine: overcoming or recapitulating structural challenges to improving patient care? *Cell Rep Med*. 2022;**3**(5):100622.
36. **Sun M, Oliwa T, Peek ME, Tung EL.** Negative patient descriptors: documenting racial bias in the electronic health record. *Health Aff (Millwood)*. 2022;**41**(2):203–211.

https://doi.org/10.1017/cts.2023.610 Published online by Cambridge University Press