ORIGINAL RESEARCH

Extracting Cognitive Impairment Assessment Information From Unstructured Notes in Electronic Health Records Using Natural Language Processing Tools: Validation with Clinical Assessment Data

Kuan-Yuan Wang[1−3],*, Mufaddal Mahesri [4],*, John Novoa-Laurentiev[4], Lily G Bessette [4], Cassandra York [4], Heidi Zakoul[4], Su Been Lee[4], Kerry Ngan[4], Li Zhou[4,5],*, Dae Hyun Kim [2,3,5],*, Kueiyu Joshua Lin[4−6],*

[1]National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [2]Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA; [3]Beth Israel Deaconess Medical Center, Boston, MA, USA; [4]Brigham and Women's Hospital, Boston, MA, USA; [5]Harvard Medical School, Boston, MA, USA; [6]Massachusetts General Hospital, Boston, MA, USA

*These authors contributed equally to this work

Correspondence: Kueiyu Joshua Lin, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont St. Suite 3030, Boston, MA, 02120, USA, Tel +1 (617) 278-0930, Fax +1 (617) 232-8602, Email jklin@bwh.harvard.edu

Purpose: We aimed to develop a Natural Language Processing (NLP) algorithm to extract cognitive scores from electronic health records (EHR) data and compare them with cognitive function recorded by Centers for Medicare & Medicaid Services (CMS)-mandated clinical assessments in nursing homes and home health visits.

Patients and Methods: We identified a cohort of Medicare beneficiaries who had either the Minimum Data Set (MDS) or Outcome and Assessment Information Set (OASIS) linked to EHR data from the Research Patient Data Registry (Mass General Brigham, Boston, MA) from 2010 to 2019. We applied an NLP approach to identify the Montreal Cognitive Assessment (MoCA) and the Mini-Mental State Examination (MMSE) scores from unstructured clinician notes in EHR. Using the NLP-extracted MoCA or MMSE scores from EHR, we compared mean differences of extracted MoCA or MMSE by cognition status determined by MDS (impaired vs intact cognition) and OASIS (severe impairment vs intact cognition) data, respectively.

Results: Our study cohort had 7419 patients who had MDS (19.7%) or OASIS (80.3%) assessments, with a mean age of 80 (SD=7) years and 60% female. In EHR, the NLP algorithm extracted cognitive test scores with 97% accuracy (95% CI: 92–99%) for MoCA and 100% accuracy (95% CI: 84–100%) for MMSE. In MDS, the mean difference in extracted MoCA was −5.6 (95% CI: −8.7, −2.4, p=0.0008), and the mean difference in extracted MMSE was −7.9 (95% CI: −12.4, −3.5, p=0.0012). In OASIS, the mean difference in extracted MoCA and extracted MMSE was −4.8 (95% CI: −9.1, −0.6, p=0.0006) and −4.5 (95% CI: −9.5, −0.5, p=0.0182), respectively.

Conclusion: We developed an NLP algorithm to accurately extract cognitive scores from unstructured EHR, and these extracted cognitive scores were well correlated with cognition function recorded in CMS-mandated clinical assessments. This could help researchers identify patients with various degrees of cognitive impairment in EHR-based research.

Keywords: natural language processing, electronic health records, cognitive impairment, mini-mental status examination, Montreal Cognitive Assessment

353

Introduction

In the United States, an estimated 6.7 million people aged 65 years and older are living with Alzheimer's dementia in 2023. As people continue to live longer, this number is projected to be as high as 13.8 million by 2060.[1] The cognitive decline encountered by patients with dementia affects their independent functioning, which subsequently leads to high healthcare utilization and the need for long-term care. As a result, a substantial financial burden is incurred on the US healthcare system, with costs estimated to be $345 billion in 2023, with the majority funded by public programs such as Medicaid and Medicare. About 48% of nursing home residents are diagnosed with Alzheimer's disease and other dementias.[1] Therefore, identification and appropriate workup for dementia is essential for chronic disease management and health-system capacity planning.

Despite the high prevalence and significant public health implications of the disease, dementia continues to be underdiagnosed, with only about half of individuals who meet the criteria being diagnosed by a clinician.[2–4] Such a condition is generally under-coded or prone to coding errors and misclassification, which may adversely affect future treatment planning and service provision. Furthermore, this also makes research in the field of Alzheimer's disease more challenging. A study found that only 11% of dementia patients had a documented cognitive measure in the five years prior to diagnosis.[5] Healthcare database research that uses diagnostic codes to identify dementia will also be limited due to the complexity of evaluating dementia in the clinical setting. Therefore, there is an unmet need to apply innovative approaches to improve the identification of patients with dementia or cognitive impairment in the healthcare data.

Patients with dementia often exhibit symptoms of cognitive impairment that are important indicators of disease status and provide insight into impairment of activities of daily living and other dementia-related outcomes. However, this information on cognitive impairments is often recorded only in free-text clinician notes within the electronic health records (EHR).[6] Since manual chart review is not scalable for large patient populations, extracting valuable information relating to patient cognition from EHR data is difficult. Because of this, several natural language processing (NLP) algorithms have been developed to help distinguish between normal cognition and any stage of cognitive impairment for patients with cognitive concerns in EHR data.[7,8] NLP-based approaches to extract test scores from clinical note text have been shown to perform well.[9,10] However, between a lack of readily available annotated data and automated tools, the application of NLP to healthcare problems presents challenges.[11] While unstructured EHR data has demonstrated its value in capturing information about dementia severity,[12] the validation of algorithms using linked claim data and Centers for Medicare & Medicaid Services (CMS)-mandated clinical assessment data is still lacking.

We aimed to develop an NLP tool to extract cognitive scores from EHR data and validate it using chart review. We also compared the NLP-extracted cognitive impairment scores with cognitive impairment assessment information in the structured data from CMS-mandated clinical assessment files for nursing homes (Minimum Data Set [MDS])[13] and home health (Outcome and Assessment Information Set [OASIS]).[14]

Materials and Methods

Database

We used CMS Medicare claims data with MDS and OASIS linked to the electronic health record (EHR) of the Mass General Brigham (MGB) system from 2010 to 2019 that includes comprehensive clinical information for over 4.7 million patients receiving care at 2 tertiary medical centers, 4 community hospitals, 3 specialty hospitals, a Rehabilitation center, and more than 35 primary care centers within the MGB system. The EHR data contain information on patient demographics, medical history, medications, vital signs, and laboratory data. The Medicare claims data include enrollment files, inpatient claims, outpatient claims, carrier claims, claims from home health agency, hospice, skilled nursing facility, and durable medical equipment, along with prescription drug information from Medicare Part D claims. MDS includes data generated from clinical assessments conducted on all residents in Medicare or Medicaid-certified nursing homes, as mandated by the federal government. It is a standardized and comprehensive assessment of a nursing home resident's functional, medical, psychosocial, and cognitive status. OASIS contains standardized data elements, including functional, medical, psychosocial, and cognitive status, collected from patients who received skilled services from home health agencies (HHA). This study was approved by the MGB Institutional Review Board, which granted an exemption

of obtaining individual informed consent. CMS granted the permission to access and utilize the study data under data use agreement between MGB and CMS.

Study Population

Patients were eligible for inclusion in the study if they were 65 years or older, with a documented admission or assessment (either quarterly or annual) in MDS, or evidence of start/resumption of care or discharge assessment in OASIS, that did not occur within 30 days of hospitalization as cognitive function can be impacted during this period.[15,16] The cohort entry (index) date was the MDS or OASIS assessment date. Patients required at least 365 days of continuous enrollment in Medicare claims data for Part A, B, and D prior to the index date and this period (365 days prior to, and including, the index date) was the covariate assessment period (CAP). We restricted the cohort to patients with a diagnosis of dementia, as recorded in the CMS Chronic Conditions Data Warehouse (CCW) dataset,[17] with at least one EHR encounter occurring in the CAP. We also excluded individuals with a diagnosis of delirium (ICD-9/ICD-10 codes in Appendix Table 1) occurring seven days prior to the index date as this may confound assessment of cognitive function.

Cognition Status and Other Study Variables

The Montreal Cognitive Assessment (MoCA)[18] and the Mini-Mental Status Examination (MMSE)[19] are commonly used psychometric tests for cognitive screening. These two cognitive tests are widely utilized in clinical practice, with established cutoffs to assess the severity of cognitive impairment. In older adults, most cases of mild cognitive impairment had MMSE scores of ≥24, with a MoCA score of 18 being comparable to an MMSE score of 24.[20] Information on MoCA and MMSE is available in clinical notes, and we used an NLP and machine learning approach to extract MoCA and MMSE scores for all patients from EHR notes. Notes included visit notes, progress notes, history and physicals, and discharge summaries documented within 180 days before the index date (to ensure temporal relevancy). MoCA and MMSE assessments from chart review were utilized as the reference standard. We had three review teams for the chart review process, each consisting of one medical doctor and a trained research assistant. Each team reviewed 150 excerpts of clinical notes of 89 patients, with 50 excerpts overlapping with the other two teams to assess internal consistency.

The MDS 3.0 used the Brief Interview for Mental Status (BIMS) to assess the cognitive function of nursing home residents. In our study, cognition was measured using a four-category variable, the cognitive function scale (CFS), based on variables available in the MDS 3.0. CFS provided a single, integrated measure of the residents' cognitive function.[21] Using this CFS variable, we classified patients as cognitively intact or mildly impaired (CFS = 1 or 2) vs moderately or severely impaired (CFS = 3 or 4). In the OASIS data, cognitive function was assessed on a 5-point Likert scale (0–4). We recategorized this to a four-category variable that defined patients having values "0 = Alert/oriented" as intact, "1 = Requires prompting" as being mildly impaired, "2 = Requires some assistance" as being moderately impaired, and "3 = Requires considerable assistance" or "4 = Totally dependent" as having severely impaired cognition status.

Demographic characteristics, including age, sex, race, language, and marital status, were obtained from EHR. Frailty was defined using a claims-based frailty index and categorized patients into robust (<0.15), prefrailty (0.15 to <0.25), mild frailty (0.25 to <0.35), and moderate-to-severe frailty (≥0.35).[22–24] We also identified seven activities of daily living (ADL) dependencies using MDS and OASIS data (Details in Appendix Table 2).

Cognitive Test Score Extractor Development

The cognitive test score extractor was coded in Python and leveraged regular expressions to extract test scores from clinical note text. Pseudocode for the score extractor tool can be found on our project Github page (https://github.com/jnlaurentiev/cog_test). The iterative development process of the score extractor is illustrated in Figure 1. A development set of 500 clinical notes were first split into sentences using the Medical Text Extraction, Reasoning, and Mapping System (MTERMS) natural language processing system.[25] Note sentences were then filtered to identify those mentioning expert-curated cognitive function test key terms (Appendix Table 3). Notes were split into sentences but retained a context window of 250 characters before/after key terms. This context window was considered to adequately capture scores corresponding to the identified cognitive test without introducing too much noise. The score extractor was
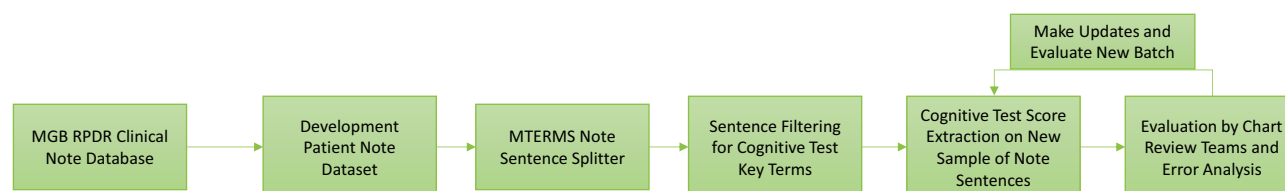
Figure 1 Iterative development process of cognitive score extractor.

iteratively developed by running the extractor on a new sample of 150 patient note sentences, reviewing extractor output for accuracy, and adjusting the regular expression code to further improve extractor performance. The NLP extractor was applied to the final cohort of 44,008 clinical notes comprising 25,512,560 excerpts to identify key terms related to cognitive function, including MMSE and the MoCA. Of these, 8159 (0.03%) excerpts from 2162 (4.9%) individual notes were found to have a cognitive test score mention.

Common sources of error were identified and mitigated over the course of development. In early rounds, the extractor sometimes struggled to distinguish between test scores and other numeric data in the notes (eg, dates/times, measures/ scores not relating to cognition). Variations in language and formatting also caused issues. For example, a MoCA score can be documented as "24/30", "24 out of 30", or just "24". Understanding how cognitive test scores are documented in the clinical notes was key to improving the performance of the score extractor.

Statistical Analysis

We calculated kappa statistics to test the inter-rater reliability for manual chart review. An ordinary least square linear regression model was used to estimate the difference with 95% confidence interval (CI) in cognitive test scores extracted from EHR between the cognitive impairment categories recorded in the MDS or OASIS data. Demographic characteristics were summarized using mean and SD for continuous variables. Additionally, we stratified our cohort by the database (MDS vs OASIS) and by the availability of cognitive scores in EHR in each database separately. We used standardized differences to quantify differences in patient characteristics across these strata in our cohort, and an absolute standardized difference of more than 0.1 was used to indicate a significant difference.[26] A two-sided p-value of less than 0.05 was considered statistically significant. All analyses were performed using SAS, version 9.4 (SAS Institute, Cary, NC).

Results

We identified 247,626 patients in MDS and 392,692 patients in OASIS data between 2010 and 2019. Among these patients, we identified our final cohort of 7419 patients, who were aged 65 years and older with a diagnosis of dementia and satisfied the inclusion and exclusion criteria (Details in Appendix Table 4). In our final cohort, 19.7% of patients had an MDS assessment (N=1462), and 80.3% had an OASIS assessment (N=5957). The mean age was 80 ± 7 years, and 25% were over age 85 years. Approximately 60% of the patients were female, 91% were non-Hispanic White, 4% were Black, 1% were Asian, and 2% were Hispanic (Table 1). The characteristics of patients from MDS and OASIS data were similar except for a few exceptions. The mean (Standard deviation [SD]) frailty score was higher among patients identified in MDS (0.32 [SD=0.1]) compared to patients in OASIS (0.25 [SD=0.1]). Patients in OASIS were also less likely to be moderate-to-severe frail compared to patients in MDS (12.9% vs 35.8%), and they also had a lower mean number of impaired baseline ADLs compared to MDS patients (4.1 ± 2.7 vs 6.1 ± 1.5).

Baseline Characteristics for Patients With a Cognitive Score (MMSE/MoCA) Compared to Those Without

In the MDS cohort, patients with an extracted MMSE/MoCA score from EHR were younger compared to patients who did not have an extracted score (78.4 ± 6.3 vs 80.1 ± 7.8). We found a similar trend in age in the OASIS cohort (78.7 ±

Table 1 Patient Characteristics – Full Cohort Stratified by MDS and OASIS Data

Covariate	Total Patients (N=7419)	(I) MDS Patients (N=1462)	(II) OASIS Patients (N=5957)	Absolute Standardized Difference*
Demographics	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	
Age, in years; Mean (SD)	80.1 (7.3)	80 (7.7)	80.11 (7.1)	0.01
Age Categories; N (%)				
65–74	1774 (24.0)	385 (26.3)	1389 (23.3)	0.07
75–85	3825 (51.6)	697 (47.7)	3128 (52.5)	0.10
> 85	1820 (24.5)	380 (26.0)	1440 (24.2)	0.04
Sex; N (%)				
Female	4407 (59.4)	885 (60.5)	3522 (59.1)	0.03
Male	3012 (40.6)	577 (39.5)	2435 (40.9)	0.03
Race; N (%)				
White	6744 (90.9)	1335 (91.3)	5409 (90.8)	0.02
Black	279 (3.8)	64 (4.4)	215 (3.6)	0.04
Asian	79 (1.1)	17 (1.2)	62 (1.0)	0.02
Hispanic	108 (1.5)	14 (1.0)	94 (1.6)	0.05
Unknown	209 (2.8)	32 (2.2)	177 (3.0)	0.05
Language; N (%)				
English speaker	6260 (84.4)	1222 (83.6)	5038 (84.6)	0.03
Non-English speaker	551 (7.4)	90 (6.2)	461 (7.8)	0.06
Unknown/missing	608 (8.2)	150 (10.3)	458 (7.7)	0.09
Marital Status; N (%)				
Married	3339 (45.0)	491 (33.6)	2848 (47.8)	0.29
Single	1013 (13.7)	292 (20.0)	721 (12.1)	0.22
Divorced/separated	623 (8.4)	181 (12.4)	442 (7.4)	0.17
Widowed	1788 (24.1)	359 (25.6)	1429 (24.0)	0.04
Unknown/other	656 (8.9)	139 (9.5)	517 (8.7)	0.03
Frailty score; Mean (SD)	0.27 (0.1)	0.32 (0.1)	0.25 (0.1)	1.27
Frailty score Categories; N (%)				
Robust (score <0.15)	216 (2.9)	<11**	214 (3.6)	0.26
Prefrail (score ≥0.15 and <0.25)	2377 (32.0)	100–150**	2256 (37.9)	0.75
Mildly frail (score ≥0.25 and <0.35)	3537 (47.7)	816 (55.8)	2721 (45.7)	0.20
Moderate-to-severely-frail (score ≥0.35)	1289 (17.4)	523 (35.8)	766 (12.9)	0.55
No. of impaired ADLs in Baseline; Mean (SD)	4.46 (2.7)	6.05 (1.5)	4.05 (2.7)	0.91

(Continued)

Table 1 (Continued).

Covariate	Total Patients (N=7419)	(I) MDS Patients (N=1462)	(II) OASIS Patients (N=5957)	Absolute Standardized Difference*
ADL Dependence; N (%)				
Groom	4748 (64.0)	1355 (92.7)	3393 (57.0)	0.90
Dress	5143 (69.3)	1374 (94.0)	3769 (63.3)	0.81
Bath	5440 (73.3)	1414 (96.8)	4026 (67.6)	0.82
Toilet	4620 (62.3)	1320 (90.3)	3300 (55.4)	0.85
Transfer	5034 (67.9)	1271 (87.0)	3763 (63.2)	0.57
Feeding	2952 (39.8)	747 (51.1)	2205 (37.0)	0.29
Ambulate	5029 (67.8)	1363 (93.2)	3666 (61.5)	0.82

Notes: *Absolute standardized Difference between (I) and (II). **Specific counts could not be provided due to the CMS cell Suppression Policy, which does not allow for cell counts between 1 and 10 to be reported.

6.7 vs 80.2 ± 7.2). In both MDS and OASIS data, a higher number of English speakers were seen among patients with an extracted score compared to those who did not have a score (MDS: 95% vs 83%; OASIS: 92% vs 84%). In the MDS cohort, patients with an extracted score had a lower mean number of impaired ADLs than patients without an extracted score (5.7 vs 6.1; absolute standardized difference = 0.24). Similarly, the mean number of impaired ADLs was noted to be lower for patients with an extracted score compared to those without an extracted score in the OASIS cohort (3.7 vs 4.1; absolute standardized difference = 0.15) (Table 2).

The Severity of Cognitive Impairment Based on MDS or OASIS

In the MDS cohort, among patients who had an extracted MoCA score (n=61), 62.3% had moderate-to-severe cognitive impairment (MoCA < 18). In contrast, in the OASIS cohort, among patients with an extracted MoCA score (N=314), 50.6% had moderate-to-severe cognitive impairment (Appendix Table 5). Among the MDS patients with MMSE score (n=25), 40% had moderate-to-severe cognitive impairment (MMSE < 24). In the OASIS patients with MMSE score (n=171), 53.2% had moderate-to-severe cognitive impairment (Appendix Table 6).

NLP Extractor Performance and the Correlation Between Extracted Scores and Clinical Assessment

In the manual chart review, three separate teams reviewed 300 excerpts from patient notes. Kappa values, used to assess inter-rater reliability, were 97.2, 94.0, and 92.9% between the three teams, respectively. This included 119 excerpts with MoCA scores and 21 with MMSE scores. In the chart review, we found that the accuracy of the NLP algorithm was 100% (95% CI, 84–100%) for MMSE and 97% (95% CI, 92–99%) for MoCA. Based on MDS data, patients identified as having impaired cognition (ie, CFS=3 or 4; mean MoCA score=11.7 [SD=4.9]) had a lower extracted MoCA score than having intact cognition (ie, CFS=1 or 2; mean MoCA score = 17.3 [SD=5.8]) with a mean difference in MoCA score of −5.58 (95% CI = −8.72, −2.43). Similarly, based on MDS data, patients identified as having impaired cognition (ie, CFS=3 or 4; mean MoCA score=16.2 [SD=4.8]) had a lower extracted MMSE score than having intact cognition (ie, CFS=1 or 2; mean MMSE score= 24.1 [SD=4.2]) with a mean difference in MMSE score of −7.95 (95% CI=−12.4, −3.49; Table 3).

In OASIS data, patients identified as having severe cognitive impairment (mean MoCA score=14.2 [SD=5.6]) had a lower extracted MoCA score than having intact cognition (mean MoCA score = 19.1 [SD=6.3]) with a mean difference in MoCA score of −4.84 (95% CI = −9.08, −0.59). Similarly, based on OASIS data, patients identified as having severe cognitive impairment (mean MMSE score=19.4 [SD=5.4]) had a lower extracted MMSE score than those having intact cognition (mean MMSE score = 23.9 [SD=6.3]) with a mean difference in MMSE score of −4.49 (95% CI=−9.50, −0.52; Table 4).

Table 2 Patient Characteristics for MDS and OASIS Cohort Stratified by Availability of Cognitive Score in EHR (MMSE or MoCA)

Covariate	MDS				OASIS			
	MDS patients (N=1462)	Patients with MOCA/MMSE (N=79)	Patients without MOCA/MMSE (N=1383)	Absolute Standardized Difference*	OASIS patients (N=5957)	Patients with MOCA/MMSE (N=439)	Patients without MOCA/MMSE (N=5518)	Absolute Standardized Difference*
Demographics	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)		Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	
Age, in years; Mean (SD)	80 (7.7)	78.4 (6.3)	80.1 (7.8)	0.24	80 (7.1)	78.7 (6.7)	80.2 (7.2)	0.22
Age Categories; N (%)								
65–74	385 (26.3)	21 (26.6)	364 (26.3)	0.01	1389 (23.3)	123 (28.1)	1266 (22.9)	0.12
75–85	697 (47.7)	44 (55.7)	653 (47.2)	0.17	3128 (52.5)	246 (56.0)	2882 (52.2)	0.08
> 85	380 (26.0)	14 (17.7)	366 (26.5)	0.21	1440 (24.2)	70 (16.0)	1370 (24.8)	0.22
Sex; N (%)								
Female	885 (60.5)	44 (55.7)	841 (60.8)	0.10	3522 (59.1)	269 (61.3)	3253 (59.0)	0.05
Male	577 (39.5)	35 (44.3)	542 (39.2)	0.10	2435 (40.9)	170 (38.7)	2265 (41.1)	0.05
Race; N (%)								
White	1335 (91.3)	76 (96.2)	1259 (91.0)	0.21	5409 (90.8)	393 (89.5)	5016 (91.0)	0.05
Black	64 (4.4)	<11**	62 (4.5)	0.11	215 (3.6)	25 (5.7)	190 (3.4)	0.11
Asian	17 (1.2)	0 (0.0)	17 (1.2)	0.16	62 (1.0)	<11**	60 (1.1)	0.07
Hispanic	14 (1.0)	0 (0.0)	14 (1.0)	0.14	94 (1.6)	<11**	89 (1.6)	0.04
Unknown	32 (2.2)	<11**	31 (2.2)	0.07	177 (3.0)	14 (3.2)	163 (3.0)	0.01
Language; N (%)								
English speaker	1222 (83.6)	75 (95.0)	1147 (83.0)	0.39	5038 (84.6)	405 (92.3)	4633 (84.0)	0.26
Non-English speaker	90 (6.2)	<11**	90 (6.5)	0.37	461 (7.7)	23 (5.2)	438 (8.0)	0.11
Unknown/missing	150 (10.3)	<11**	146 (10.6)	0.21	458 (7.7)	11 (2.5)	447 (8.1)	0.25

(*Continued*)

Table 2 (Continued).

Covariate	MDS				OASIS			
	MDS patients (N=1462)	Patients with MOCA/MMSE (N=79)	Patients without MOCA/MMSE (N=1383)	Absolute Standardized Difference*	OASIS patients (N=5957)	Patients with MOCA/MMSE (N=439)	Patients without MOCA/MMSE (N=5518)	Absolute Standardized Difference*
Marital Status; N (%)								
Married	491 (33.6)	27 (34.2)	464 (33.6)	0.01	2848 (47.8)	188 (42.8)	2660 (48.2)	0.11
Single	292 (20.0)	15 (19.0)	277 (20.0)	0.03	721 (12.1)	63 (14.4)	658 (11.9)	0.07
Divorced/separated	181 (12.4)	15 (19.0)	166 (12.0)	0.19	442 (7.4)	56 (12.8)	386 (7.0)	0.19
Widowed	359 (25.6)	10–30**	343 (24.8)	0.11	1429 (24.0)	119 (27.1)	1310 (23.7)	0.08
Unknown/other	139 (9.5)	<11**	133 (9.6)	0.07	517 (8.7)	13 (3.0)	504 (9.1)	0.26
Frailty score; Mean (SD)	0.33 (0.1)	0.34 (0.1)	0.33 (0.1)	0.15	0.27 (0.1)	0.27 (0.1)	0.27 (0.1)	0.00
Frailty score Categories; N (%)								
Robust (score <0.15)	<11**	<11**	<11**	0.05	214 (3.6)	11 (2.5)	203 (3.7)	0.07
Prefrail (score >0.15 and <0.25)	100–150**	<11**	100–150**	0.08	2256 (37.9)	160 (36.5)	2096 (38.0)	0.03
Mildly frail (score ≥0.25 and <0.35)	816 (55.8)	40 (50.6)	776 (56.1)	0.11	2721 (45.7)	206 (47.0)	2515 (45.6)	0.03
Moderate-to-severely-frail (score >0.35)	523 (35.8)	34 (43.0)	489 (35.4)	0.16	766 (12.9)	62 (14.1)	704 (12.8)	0.04
No. of impaired ADLs in Baseline; Mean (SD)	6.05 (1.5)	5.7 (1.7)	6.07 (1.4)	0.24	4.05 (2.7)	3.68 (2.7)	4.08 (2.7)	0.15
ADL Dependence; N (%)								
Groom	1355 (92.7)	68 (86.1)	1287 (93.1)	0.23	3393 (57.0)	229 (52.2)	3164 (57.3)	0.10
Dress	1374 (94.0)	71 (89.9)	1303 (94.2)	0.16	3769 (63.3)	257 (58.6)	3512 (63.7)	0.10

Bath	1414 (96.7)	75 (94.9)	1339 (96.8)	0.09	4026 (67.6)	290 (66.1)	3736 (67.7)	0.04
Toilet	1320 (90.3)	67 (84.8)	1253 (90.6)	0.18	3300 (55.4)	223 (50.8)	3077 (55.8)	0.10
Transfer	1271 (87.0)	67 (84.8)	1204 (87.1)	0.06	3763 (63.2)	247 (56.3)	3516 (63.7)	0.15
Feeding	747 (51.1)	30 (38.0)	717 (51.8)	0.28	2205 (37.0)	124 (28.3)	2081 (37.7)	0.20
Ambulate	1363 (93.2)	71 (89.9)	1292 (93.4)	0.13	3666 (61.5)	247 (56.3)	3419 (62.0)	0.12

Notes: *Absolute standardized Difference between (I) and (II). **Specific counts could not be provided due to the CMS cell Suppression Policy, which does not allow for cell counts between 1 and 10 to be reported.

Table 3 Association of MoCA and MMSE Score in EHR with Cognitive Status as Defined in MDS Data

Cognition in MDS	Extracted MoCA Score; Mean (SD)	Difference (95% CI)	Extracted MMSE Score; Mean (SD)	Difference (95% CI)
A) Intact	17.3 (5.8)	Reference	24.1 (4.2)	Reference
B) Impaired	11.7 (4.9)	−5.58 (−8.72, −2.43)	16.2 (4.8)	−7.95 (−12.41, −3.49)

Notes: A) Intact in MDS= CFS 1 or 2; B) Impaired in MDS= CFS 3 or 4.

Table 4 Association of MoCA and MMSE Score in EHR with Cognitive Status as Defined in OASIS Data

Cognition in OASIS	Extracted MoCA Score; Mean (SD)	Difference (95% CI)	Extracted MMSE Score; Mean (SD)	Difference (95% CI)
A) Intact	19.1 (6.3)	Reference	23.9 (6.3)	Reference
B) Mild impairment	16.9 (5.4)	−2.15 (−3.68, −0.62)	20.4 (6.9)	−3.46 (−5.96, −0.96)
C) Moderate impairment	15.7 (6.0)	−3.36 (−5.10, −1.62)	20.1 (6.6)	−3.78 (−6.48, −1.09)
D) Severe impairment	14.2 (5.6)	−4.84 (−9.08, −0.59)	19.4 (5.4)	−4.49 (−9.50, −0.52)

Notes: A) Intact in OASIS: M1700 = 00; B) Mild impairment in OASIS: M1700 = 01; C) Moderate impairment in OASIS: M1700 = 02; D) Severe impairment in OASIS: M1700 = 03 or 04.

Discussion

We have developed and validated an NLP tool that can accurately extract cognitive scores from free-text EHR (accuracy>97%). We found that NLP-extracted cognitive impairment scores from EHR were well correlated with the cognitive impairment status in the CMS-mandated clinical assessments from nursing homes and home care agencies. In our study cohort, among those with MDS or OASIS data available, only a small proportion (5–7%) of the patients have their cognitive score recorded in free-text EHR, and their characteristics are appreciably different from those without cognitive scores recorded in the EHR.

Our findings about the change in cohort size and patient characteristics when different levels of data availability are required can help inform future study designs using local EHR or CMS datasets (MDS or OASIS). Among those with either MDS or OASIS data available, we found that the cohort size was reduced by >90% when we required patients to have cognitive scores (MMSE or MoCA) available in the EHR. Notably, those with extracted MMSE/MoCA scores were younger and had fewer impaired ADLs. English proficiency was also higher in those with than without extracted scores, which can possibly be explained by the challenges in assessing cognitive function in patients with language barriers. While having the cognitive score provides much more granular data on a continuous scale for cognitive impairment, restriction to those with such information may affect the generalizability of the study findings (ie, the results were drawn from a non-representative sample). Our study provides solid data to inform the trade-off between internal validity (access to more detailed outcomes or covariate information like MMSE/MoCA scores) and external validity (eg, generalizability) during the design stage. Both MDS and OASIS are CMS-mandated clinical assessments that are updated quarterly, provided the patients are in a nursing facility (for MDS) or receiving home health care, including skilled nursing, physical therapy, occupational therapy, or aide services (for OASIS). While we demonstrated that the extracted MMSE/MoCA is well correlated with the cognitive assessment recorded in MDS and OASIS, MDS and OASIS generally provide a more reliable and structured longitudinal trajectory of cognitive function over time given that it is required for CMS reimbursement for the respective services. Older adults diagnosed with dementia often undergo frequent transitions between healthcare settings,[27] requiring continuous monitoring and coordinated care to ensure optimal management of their condition. The severity of dementia plays a vital role in enabling healthcare providers to prescribe appropriate treatments and connect patients and caregivers with essential resources. Our approach provides readily accessible disease

severity information across healthcare settings, supporting timely diagnosis and early intervention to manage dementia progression effectively.

Although the documentation of cognitive tests in EHRs seems limited, real-world data remain valuable for dementia research.[28] Extracting different cognitive tests and tracking cognitive impairment severity can benefit EHR-based dementia studies. We have developed the NLP tool to extract the most commonly used cognitive scores in clinical practice (ie, MMSE and MoCA). Our tool has an excellent accuracy (>97%) compared to manual chart review results. In the literature, high accuracy rates have been observed for NLP tools to extract other clinical factors or scores. For instance, Wagholikar et al developed an extractor for ejection fractions with ~100% accuracy rate when validated manually by cardiologists.[29] Adekkanattu et al developed a tool to extract Patient Health Questionnaire–9 (PHQ-9) scores with an accuracy rate of 97%.[30] These studies collectively underscore the reliability of NLP tools in extracting numerical values for a clinical score or test results with an unambiguous name for the instrument. A prior study employed a rule-based approach to extract information from unstructured clinicians' notes to identify dementia severity.[12] However, this study included patients aged 40 years or older and relied on cognitive scores and explicit terms from a single healthcare system to determine dementia severity. In contrast, our study extracted MMSE and MoCA cognitive scores and separately compared them to CMS-mandated clinical assessments. We compared the characteristics of patients from MDS and OASIS datasets to identify differences and similarities, offering insights into patient demographics, frailty, and functional status. Furthermore, we categorized cognitive test scores into severity categories to track dementia progression over time before older patient transition to nursing homes or home care. Cognitive test scores were extracted from EHR notes within 180 days prior to the MDS or OASIS assessment date. In order to extract the most recent and clinically relevant cognitive test score, we did not include patient notes older than 6 months. In the context of cognitive impairment, we felt that this time window was clinically meaningful in offering valuable insights into disease progression and care needs over time. Detecting evidence of cognitive decline from clinical notes can complement structured EHR data, providing a more comprehensive assessment of a patient's cognitive status.

Our study has several limitations. First, our findings were developed from a metropolitan academic system and may not be generalizable to different settings. In the United States, approximately 46% of nursing home residents have Alzheimer's disease or another form of dementia.[31] In our study, although the proportion of patients with extracted cognitive scores was low, 40–62.3% of those in the MDS cohort exhibited moderate-to-severe cognitive impairment. Therefore, our study cohort may not fully represent the broader population of nursing home residents with dementia. Rule-based algorithms can be tailored to meet the specific needs of healthcare systems and research studies, allowing for flexibility in identifying, classifying, and analyzing clinical data based on institutional requirements and study objectives. Second, MDS and OASIS data are not collected specifically for research purposes, although they are considered reliable and valid tools for population health studies.[32,33] As is the case with administrative claims data, there remains a possibility of misclassification for some clinical variables. Our study focused on MMSE and MoCA extraction as these are the most commonly used cognitive scores in clinical practice.[34,35] Moreover, the MMSE is useful for assessing cognitive functions, but the test alone cannot be used to diagnose dementia.[36] On the other hand, MoCA was originally developed to detect mild cognitive impairment and is now frequently used as a screening tool for dementia.[37] Our study did not aim to differentiate mild cognitive impairment from intact cognition. Rather, we sought to develop and validate a tool to extract cognitive scores that can be used to track cognitive decline over time. From a clinical perspective, understanding residents' cognitive decline allows for the targeted allocation of resources and implementation of interventions to address modifiable factors and help mitigate further decline. However, older patients with low English literacy and limited education may perform poorly. This may explain why the patient characteristics differ between those with MMSE/MoCA and those without the tests. The former subjects were younger, less functionally dependent, and almost all were English speakers. Furthermore, over 90% of our study patients are white, which is considered a lack of racial/ethnic diversity. Lastly, we would like to acknowledge that our study includes EHR notes occurring over 10 years (2010–2019) and as health record documentation practices change over time there may be some potential differences based on calendar years.

Our approach has successfully made previously inaccessible information readily available to clinicians. To ensure broader applicability, efforts should be made to validate our approach across diverse patient populations and healthcare settings. Furthermore, integrating NLP-driven automation into EHR systems could facilitate real-time data extraction and

significantly enhance clinical decision-making and patient care. In conclusion, we have developed a natural language processing algorithm that is able to extract cognitive scores from unstructured patient notes in EHR with high accuracy rates (MoCA: 97%, MMSE: 100%) when validated against chart review. The extracted cognitive scores correlate well with the structured clinical assessment. This tool can help with the identification of patients with various degrees of cognitive impairment in future research based on EHR.

Disclosure

DHK is supported by the grants R01AG071809 and K24AG073527 from the National Institute on Aging of the National Institutes of Health for unrelated work. He received a personal fee from Alosa Health and VillageMD for unrelated work. All other authors have no conflicts to declare concerning this work.

References

1. Magaziner J, German P, Zimmerman SI, et al. The prevalence of dementia in a statewide sample of new nursing home admissions aged 65 and older: diagnosis by expert panel. Epidemiology of dementia in nursing homes research group. *Gerontologist*. 2000;40(6):663–672. doi:10.1093/geront/40.6.663

2. Alzheimer's Association. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2022;18(4):700–789. doi:10.1002/alz.12638

3. Alzheimer's Association. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2023;19(4):1598–1695. doi:10.1002/alz.13016

4. Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of dementia: an observational study of patterns in diagnosis and awareness in US older adults. *J Gen Intern Med*. 2018;33(7):1131–1138. doi:10.1007/s11606-018-4377-y

5. Maserejian N, Krzywy H, Eaton S, Galvin JE. Cognitive measures lacking in EHR prior to dementia or Alzheimer's disease diagnosis. *Alzheimers Dement*. 2021;17(7):1231–1243. doi:10.1002/alz.12280

6. Kharrazi H, Anzaldi LJ, Hernandez L, et al. The value of unstructured electronic health record data in geriatric syndrome case identification. *J Am Geriatr Soc*. 2018;66(8):1499–1507. doi:10.1111/jgs.15411

7. Maclagan LC, Abdalla M, Harris DA, et al. Can patients with dementia be identified in primary care electronic medical records using natural language processing? *J Healthc Inform Res*. 2023;7(1):42–58. doi:10.1007/s41666-023-00125-6

8. Noori A, Magdamo C, Liu X, et al. Development and evaluation of a natural language processing annotation tool to facilitate phenotyping of cognitive status in electronic health records: diagnostic study. *J Med Internet Res*. 2022;24(8):e40384. doi:10.2196/40384

9. Gandomi A, Hasan E, Chusid J, et al. Evaluating the accuracy of lung-RADS score extraction from radiology reports: manual entry versus natural language processing. *Int J Med Inform*. 2024;191:105580. doi:10.1016/j.ijmedinf.2024.105580

10. Yu S, Le A, Feld E, et al. A natural language processing-assisted extraction system for Gleason scores: Development and usability study. *JMIR Cancer*. 2021;7(3):e27970. doi:10.2196/27970

11. Hossain E, Rana R, Higgins N, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. *Comput Biol Med*. 2023;155:106649. doi:10.1016/j.compbiomed.2023.106649

12. Prakash R, Dupre ME, Østbye T, Xu H. Extracting critical information from unstructured clinicians' notes data to identify dementia severity using a rule-based approach: feasibility study. *JMIR Aging*. 2024;7:e57926. doi:10.2196/57926

13. Services. RDACCfMM. Long Term Care Minimum Data Set (MDS) 3.0. Available from: https://resdac.org/cms-data/files/mds-30. Accessed April 4, 2025.

14. Services. RDACCfMM. *Home Health Outcome and Assessment Information Set*. Available from https://resdac.org/cms-data/files/oasis. Accessed, 2025.

15. Lindquist LA, Go L, Fleisher J, Jain N, Baker D. Improvements in cognition following hospital discharge of community dwelling seniors. *J Gen Intern Med*. 2011;26(7):765–770. doi:10.1007/s11606-011-1681-1

16. Saczynski JS, McManus DD, Waring ME, et al. Change in cognitive function in the month after hospitalization for acute coronary syndromes: findings from TRACE-CORE (Transition, risks, and actions in coronary events-center for outcomes research and education). *Circ Cardiovasc Qual Outcomes*. 2017;10(12). doi:10.1161/circoutcomes.115.001669

17. (CMS) CfMMS. Chronic conditions data warehouse.

18. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695–699. doi:10.1111/j.1532-5415.2005.53221.x

19. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–198. doi:10.1016/0022-3956(75)90026-6

20. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin AJ. Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. *BMC Geriatr*. 2015;15:107. doi:10.1186/s12877-015-0103-3

21. Thomas KS, Dosa D, Wysocki A, Mor V. The minimum data set 3.0 cognitive function scale. *Med Care*. 2017;55(9):e68–e72. doi:10.1097/mlr.0000000000000334

22. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol a Biol Sci Med Sci*. 2007;62(7):722–727. doi:10.1093/gerona/62.7.722

23. Shi SM, McCarthy EP, Mitchell S, Kim DH. Changes in predictive performance of a frailty index with availability of clinical domains. *J Am Geriatr Soc*. 2020;68(8):1771–1777. doi:10.1111/jgs.16436

24. Shi SM, McCarthy EP, Mitchell SL, Kim DH. Predicting mortality and adverse outcomes: comparing the frailty index to general prognostic indices. *J Gen Intern Med*. 2020;35:1516–1522. doi:10.1007/s11606-020-05700-w

25. Zhou L, Plasek JM, Mahoney LM, et al. Using medical text extraction, reasoning and mapping system (MTERMS) to process medication information in outpatient clinical notes. *AMIA Annu Symp Proc*. 2011;2011:1639–1648.

26. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083–3107. doi:10.1002/sim.3697

27. Callahan CM, Tu W, Unroe KT, LaMantia MA, Stump TE, Clark DO. Transitions in care in a nationally representative sample of older Americans with dementia. *J Am Geriatr Soc*. 2015;63(8):1495–1502. doi:10.1111/jgs.13540

28. Chen Z, Zhang H, Yang X, et al. Assess the documentation of cognitive tests and biomarkers in electronic health records via natural language processing for Alzheimer's disease and related dementias. *Int J Med Inform*. 2023;170:104973. doi:10.1016/j.ijmedinf.2022.104973

29. Wagholikar KB, Fischer CM, Goodson A, et al. Extraction of ejection fraction from echocardiography notes for constructing a cohort of patients having heart failure with reduced ejection fraction (HFrEF). *J Med Syst*. 2018;42(11):209. doi:10.1007/s10916-018-1066-7

30. Adekkanattu P, Sholle ET, DeFerio J, Pathak J, Johnson SB, Campion TR Jr. Ascertaining depression severity by extracting patient health questionnaire-9 (PHQ-9) scores from clinical notes. *AMIA Annu Symp Proc*. 2018;2018:147–156.

31. Alzheimer's Association. 2024 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2024;20(5):3708–3821. doi:10.1002/alz.13809

32. Mor V, Intrator O, Unruh MA, Cai S. Temporal and geographic variation in the validity and internal consistency of the nursing home resident assessment minimum data set 2.0. *BMC Health Serv Res*. 2011;11:78. doi:10.1186/1472-6963-11-78

33. Tullai-McGuinness S, Madigan EA, Fortinsky RH. Validity testing the outcomes and assessment information set (OASIS). *Home Health Care Serv Q*. 2009;28(1):45–57. doi:10.1080/01621420802716206

34. Damian AM, Jacobson SA, Hentz JG, et al. The Montreal Cognitive Assessment and the Mini-mental State Examination as screening instruments for cognitive impairment: item analyses and threshold scores. *Dement Geriatr Cogn Disord*. 2011;31(2):126–131. doi:10.1159/000323867

35. Young J, Meagher D, Maclullich A. Cognitive assessment of older people. *BMJ*. 2011; 343:d5042. doi:10.1136/bmj.d5042

36. Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-mental state examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database Syst Rev*. 2016;2016(1):Cd011145. doi:10.1002/14651858.CD011145.pub2

37. Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal cognitive assessment for the detection of dementia. *Cochrane Database Syst Rev*. 2021;7(7):Cd010775. doi:10.1002/14651858.CD010775.pub3