



Published in final edited form as:

Alzheimers Dement. 2020 March ; 16(3): 531–540. doi:10.1016/j.jalz.2019.09.084.

Stratifying risk for dementia onset using large-scale electronic health record data: a retrospective cohort study

Thomas H. McCoy Jr., MD^{*1}, Larry Han, MPhil², Amelia M. Pellegrini, BA³, Rudolph E. Tanzi, PhD⁴, Sabina Berretta, MD⁵, Roy H. Perlis, MD MSc³

¹Center for Quantitative Health, Massachusetts General Hospital, Boston, MA, USA

²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Center for Quantitative Health, Massachusetts General Hospital, Boston, MA, USA

⁴Genetics and Aging Research Unit, McCance Center for Brain Health, MassGeneral Institute for Neurodegenerative Disease, Department of Neurology, Massachusetts General Hospital, Charlestown, MA, USA

⁵Translational Neuroscience Lab., Basic Neuroscience Division, McLean Hospital, Belmont, MA, USA

Abstract

INTRODUCTION: Preventing dementia, or modifying disease course, requires identification of pre-symptomatic or minimally symptomatic high-risk individuals.

METHODS: We utilized longitudinal electronic health records from two large academic medical centers and applied a validated natural language processing tool to estimate cognitive symptomatology. We used survival analysis to examine the association of cognitive symptoms with incident dementia diagnosis during up to 8 years of follow-up.

RESULTS: Among 267,855 hospitalized patients with 1,251,858 patient-years of follow-up data, 6,516 (2.4%) received a new diagnosis of dementia. In competing risk regression, increasing cognitive symptom score was associated with earlier dementia diagnosis (HR 1.63; 1.54 – 1.72). Similar results were observed in the second hospital system, and in subgroup analysis of younger and older patients.

DISCUSSION: A cognitive symptom measure identified in discharge notes facilitated stratification of risk for dementia up to 8 years before diagnosis.

*Correspondence: Thomas H. McCoy, Jr. MD, Massachusetts General Hospital, 185 Cambridge Street, 6th Floor, Boston, MA, 02114 USA, Ph 617-643-6310, Fax 617-726-7541, thmccoy@partners.org.

¹Abbreviations: EHR, electronic health record; NLP, natural language processing

Authorship

THM revised the manuscript, planned the experiments, developed the original software, and analyzed data. LH assisted with statistical analysis and interpretation. AMP assisted in manuscript revision and submission. RET and SB drafted and revised the manuscript and provided input on experimental design. RHP drafted and revised the manuscript, planned the experiments, oversaw data collection and characterization, and advised on analysis.

Keywords

Alzheimer's disease; dementia; cognition; natural language processing; machine learning; electronic health record; phenotype; Research Domain Criteria; data mining

1. INTRODUCTION¹

While currently marketed treatments for dementia exhibit modest efficacy overall, they are thought to be most effective when used early in the disease course.[1–3] As such, early detection of those who will develop dementia is a public health priority.[4]

To date, the only validated *in vivo* markers of early or preclinical disease rely on PET imaging, although such imaging is not definitive in diagnosis[5,6] and widespread clinical deployment of PET is not feasible at present. In a research context, minimally burdensome risk-enriched recruitment, particularly by primary care doctors, is a recognized essential component of the National Plan to Address Alzheimer's Disease.[7] Challenges in study recruitment in this context have been implicated in past failures to bring new compounds to market.[8] Although numerous cognitive screening measures for dementia have been developed, their reliability and scalability are variable.[9,10] As such, additional low-cost high-scale means of dementia risk stratification would be useful in prioritizing next-step assessment in clinical care and facilitating clinical research on early intervention. The aim of this report is to assess an electronic health record (EHR)-based screening approach that could help to meet this need.

In particular, we sought to determine whether concepts derived from narrative hospital discharge notes enable risk stratification of future dementia diagnosis. Narrative free text is widely available in EHRs but unwieldy to work with in conventional analytic paradigms; as such, tools for quantifying concepts within these documents for subsequent analysis are an area of active research.[11] We have recently demonstrated that natural language processing (NLP) applied to EHR documentation may capture useful elements of clinical presentation not necessarily reflected in coded data. For example, non-coded psychiatric symptoms captured in clinical documentation improve prediction of suicide risk,[12,13] the subjective valence of notes predict all-cause mortality,[14] and symptoms or symptom domains predict future utilization.[15,16] To extend the literature on the utility of NLP in the clinical prediction of dementia risk stratification, we applied a recently-validated approach to estimating cognitive symptoms in narrative notes from two large academic medical centers with distinct clinical populations.[17,18] To evaluate the derived cognitive symptom scores we examined the extent to which cognitive score associated with subsequent dementia risk among hospitalized individuals without a prior diagnosis of dementia. Secondly, we examined the association between this score and death among hospitalized individuals already diagnosed with dementia, hypothesizing that a common set of features might also represent a marker of disease progression and severity, shown to associate with mortality. [19,20]

2. MATERIALS AND METHODS

2.1 Study design and cohort derivation

This study used a retrospective cohort design drawing on inpatient discharges documented in the EHR of two large academic medical centers for adults (age 18 years and older) hospitalized between January 1, 2005 and December 31, 2017. In addition to sociodemographic features (age at admission, sex, and race/ethnicity), we extracted administrative diagnostic codes and narrative hospital discharge summary notes. The *primary* study cohort (used to examine dementia onset, referred to subsequently as the incident dementia cohort) included all individuals who had not received a structured diagnosis of dementia during the index hospitalization, or at any point prior to that hospitalization in either hospital's EHR. After primary analysis of the sites individually, their data were pooled for a merged analysis. The *secondary* study cohort (used to examine mortality, referred to subsequently as the prevalent dementia cohort) included individuals who received a structured diagnosis of dementia at or preceding the index hospitalization. That is, individuals with and without dementia at or before index hospitalization were separated into distinct cohorts, to allow investigation of emergence of dementia in those not previously diagnosed, and then progression of dementia to death among those who had been previously diagnosed. Diagnosis was ascertained using clinical billing data in terms of International Statistical Classification of Diseases and Related Health Problems versions 9 and 10 codes as well as structured problem list codes for dementia.[21–24]

In this context, index hospitalization refers either to the only hospitalization during the risk period, or for individuals with multiple hospitalizations, to a randomly-selected hospitalization. Random selection is consistent with prior work and avoids the bias of selecting the first (or last) observed hospitalization from the observation period at the expense of total follow-up duration.[12] (See section 2.4, below, for description of sensitivity analyses incorporating all admissions).

Datamart generation used Informatics for Integrating Biology and the Bedside, or i2b2, server software.[25,26] The Partners HealthCare Human Research Committee reviewed and approved the study protocol, waiving the requirement for informed consent as a retrospective health care utilization study with no participant contact.

2.2 Scoring clinical text for cognition symptom burden

We have previously described derivation and validation of a method for estimating neuropsychiatric symptom domains from narrative text.[17] In brief, this method relies on recognizing a prespecified set of terms within a given symptom domain, cognition in this case. This term list was developed in an iterative process which began with consensus term lists developed by a group of clinical experts, including the NIMH Research Domain Criteria working group, which were subsequently expanded through unsupervised machine learning to enhance coverage of the clinical lexicon. The final symptom score is simply the percent of terms that appear in any given note: for example, if there are 20 possible cognition-related terms, and 5 appear in a note, the note would be assigned a score of 5/20, or 25%. The scoring approach is implemented as freely available software for inspection and

the full list of tokens is available online at <https://github.com/thmccoy/CQH-Dimensional-Phenotyper> and described in the initial validation publication.[17] Of note, this tool was not developed or trained to predict emergence of dementia in any way, but simply to capture dimensions of neuropsychiatric symptoms reflecting interest in conceptualizing illness as dimensional rather than categorical.[27,28]

2.3 Outcome

The primary study outcome was time to receipt of a diagnostic code of dementia at any inpatient or outpatient visit, among individuals who had not received such a code at or prior to the index hospitalization (i.e., incident dementia). In the second, distinct cohort composed of those with a diagnosis of dementia at or before the index hospitalization (i.e., prevalent dementia), the outcome was time to death. For the secondary study in prevalent dementia cases, all-cause mortality was determined by integrating federal Social Security registry data with public death certificate data available from the Massachusetts Department of Public Health.

2.4 Analysis

In primary analysis, we examined time from hospital discharge to incident dementia diagnosis, censoring at end of follow-up, death, or 8 years, whichever came first, using survival methodology. As death precludes the incident diagnosis of dementia we approached death a competing risk and analyzed the association between cognitive symptom burden score and time to incident dementia diagnosis in a Fine-Gray subdistribution hazard model adjusting for age, sex, white race, and Charlson comorbidity index.[29–31] This analysis used the package *fastcmprsk* v1.02.[32] In this analysis continuous covariate were included directly. In keeping with prior work, we multiple the cognitive burden score by 10 to simplify communication and reduce the size of model coefficients.[33–35] To visualize censored data we plotted Kaplan-Meier curves by quartile of symptom burden score using *ggplot2* v3.1 and tested time to event differences among these groups using both univariate and stratified analysis with *survival* v2.43.[36] Where used, the stratified Kaplan-Meier log-rank tests were stratified by cognitive burden quartile, age and Charlson comorbidity index tertile, white race, and sex. To assess the performance of the cognitive symptom burden score as a classifier of censored data we report the C index of the symptom score with respect to the dementia outcome using *survival* v2.43.[37–39] To further characterize the magnitude of risk associated with the cognitive symptom score captured by the Fine-Gray analysis and depicted in Kaplan-Meier curves in a more familiar cross sectional paradigm, we created sub-cohorts including only those individuals with adequate follow up time to observe dementia at five years and evaluated dementia within five years as a binary outcome. In secondary analyses of these sub-cohorts with follow up sufficient to study a five-year dementia outcome, we report the area under the receiver operating characteristics curve (AUC) for cognitive symptom burden as a univariate classifier.[40,41] AUCs were calculated using the *pROC* v1.15 package. In the same 5-year follow up sub-cohorts we applied Youden's method to select the optimal cut point in the cognitive symptom burden score for identification of five-year dementia risk and then calculated sensitivity and specificity based on this cut point.[42,43] Cut point analysis used *cutpointr* v0.7.6. Finally, we calculated lift – the rate of outcome in a strata relative to that in the group overall – by quartile of cognitive

symptom burden score for each cohort.[44] All analysis for the primary aim of assessing association between cognitive symptom scale and new dementia diagnosis was performed for each site cohort individually and then with the two pooled. Thereafter the pooled cohort was used for a range of supplemental and sensitivity analyses which are reported primarily in the supplemental materials. As an assessment of age effects, the pooled sample was stratified into an older cohort (age 50 and greater) and a younger cohort (those less than 50 years old) and analyzed as censored survival data. Patients who were primarily admitted for stroke were sub analyzed in a survival paradigm. The five-year sub-cohort of the two-site pooled cohort was then analyzed using logistic regression minimally adjusting for age, sex, race, and Charlson for comparison in sensitivity analysis allowing multiple encounters per patient. excluding diagnosis within two years of index hospitalization, adding primary discharge diagnostic information categorized based on Agency for Healthcare Research and Quality Clinical Classification Software ICD10 level 1 codes, primary diagnosis of stroke, and finally in models stratified by quartile of age.

To better understand the timing of incident dementia diagnosis, we capitalized on the inspectability of the pre-specified term list approach and evaluated the impact of the individual terms contributing to the cognitive symptom score. To do so, we divided cases of incident dementia into near term (initial diagnosis within 1 year) and long term (initial diagnosis greater than 1 year after index hospitalization) and compared the proportion of discharges with each term present using chi squared testing and crude odds ratio. Although this analysis decomposes the previously validated composite score, it provides visibility into the potential clinical features associated with outcomes. This analysis was limited to those tokens from the cognition score that occurred in at least 10 cases of near-term and long-term dementia.

In a second and distinct cohort, as a secondary aim, we sought to understand the relationship between the cognitive symptom burden and mortality risk. In this prevalent dementia cohort (i.e., individuals with an existing dementia diagnosis at the randomly index hospitalization used for all primary analysis), we repeated stratified Kaplan-Meier analysis as described in the description of the primary aim methods above using all-cause mortality over the 8-year follow up period as the lone endpoint. This analysis gives insight into the potential for the computed cognitive symptom burden score to stratify the risk of progression from diagnosis to death, whereas the primary analysis focuses on the risk of progression to dementia diagnosis.

All analyses utilized R v3.5.2 and the specific packages described above.[45] As we tested two independent hypotheses (association with incident dementia, and then association with mortality), uncorrected p-values are reported with alpha set at 0.05. In the exploratory analysis of individual terms among the cases of incident dementia, Bonferroni-corrected p-values were used to control type 1 error.

3. RESULTS

In the first hospital system, there were 267,855 index hospitalizations of individuals without a prior dementia diagnosis and no dementia diagnosis during hospitalization, with 1,251,858

patient-years of follow-up data (Table 1). Among them, 6,516 (2.4%) received a diagnosis of dementia during up to 8 years of follow-up. In competing risk regression, cognitive score at discharge was positively associated with subsequent dementia diagnosis (aHR 1.63, 95% CI 1.54 – 1.72; Table 2A). In Kaplan-Meier survival models (Figure 1A), quartile of cognitive score was associated with emergence of dementia (log-rank chi square (3df)=570; $p < 1e-6$); significant association persisted in models stratified for age, sex, white race, and Charlson comorbidity score (log-rank chi square (3df)=309; $p < 1e-6$). The C index for cognitive burden score in the primary cohort was 0.61 (95% CI 0.60 – 0.61). In the primary cohort, 99th percentile dementia-free survival (i.e., time to 1% incidence of dementia) occurred at 1,160 days (95% CI 1041–1289 days) in the lowest quartile of cognitive symptom burden versus 189 days (95% CI 161–229 days) in the highest symptom burden quartile.

Next, we examined the generalizability of this cognitive dimension to a second academic medical center (Table 1). For 267,959 hospitalizations of individuals without a prior diagnosis of dementia, 3,902 (1.5%) subsequently received this diagnosis during 1,319,935 patient-years of follow-up data. As in the first hospital, cognitive risk score was positively associated with dementia risk in competing risk regression (aHR 1.37, 95% CI 1.29 – 1.46; Table 2B). Statistically significant differences were likewise observed using unstratified analysis of Kaplan-Meier survival curves (log-rank chi square (3df)=879; $p < 1e-6$; Figure 1B) and in stratified analysis (log-rank chi square (3df)=232; $p < 1e-6$). The C index for cognitive burden score predicting dementia in the second hospital was 0.65 (95% CI 0.64 – 0.66). In this site's cohort, 99th percentile dementia-free survival occurred at 2,186 days (95% CI 2037–2356 days) in the lowest quartile of cognitive symptom burden versus 347 days (95% CI 298–399 days) in the highest symptom burden quartile.

In follow-up pooled analysis over the two hospital systems that included 535,814 hospitalized patients, 10,418 (1.9%) of whom went on to an eventual dementia diagnosis, spanning 2,571,793 patient-years of follow-up data, the cognitive symptom burden score was significantly associated with hazard for dementia diagnosis (aHR 1.5; 95% CI 1.44 – 1.56) in competing risk regression (eTable 1) and cognitive symptom burden score by quartile in both stratified (log-rank chi squared (3df)=429; $p < 1e-6$) and unstratified (log-rank chi squared (3df)=1464; $p < 1e-6$) log rank tests (eFigure 1). The C index in the pooled cohort was 0.62 [0.62 – 0.63]. In the pooled analysis, 99th percentile dementia-free survival occurred at 1,618 days (95% CI 1519–1772 days) in the lowest quartile of cognitive symptom burden versus 257 days (95% CI 230–293 days) in the highest symptom burden quartile.

In a subgroup analysis splitting the pooled cohort at 50 years of age, cognitive burden score remained associated with risk of dementia in both groups. Among older patients ($n = 307, 525$) with a median follow up of 1,565 days higher quartile of cognitive symptom score was associated with earlier diagnosis of dementia (log-rank chi squared (3df)=430; $p < 1e-6$; eFigure 2a). The same pattern was observed in younger patients ($n = 228,289$) with higher cognitive score associated with earlier diagnosis (log-rank chi squared (3df)= 109; $p < 1e-6$; eFigure 2b) over a median follow up of 2,392 days. The older sub-group C index was 0.58 (95% CI 0.58 – 0.59) in whereas the younger sub-group C index was 0.71 (95% CI 0.68 – 0.74). In sub-group analysis limiting to those patients who had a primary discharge

diagnosis of stroke (n = 18,681) from their index hospitalization the C index for cognitive symptom score was 0.59 (95% CI 0.57 – 0.61) and higher cognitive symptom score quartile was associated with earlier dementia diagnosis in stratified analysis (log-rank chi squared (3df)=45; p<1e-6; eFigure 3) over a median follow up of 1,779 days.

To estimate size of stratification effect, an indication of the capacity of this measure to enrich for a high-risk patient subset, we examined cumulative incident of dementia at five years, by quartile of index hospitalization cognitive symptom burden, as compared to the overall rate both per hospital and pooled (eFigure 4). In the pooled cohort, individuals in the highest-risk quartile were three times more likely to receive a diagnosis of dementia than the cohort overall (that is, the ‘lift’ afforded by this model was 3). In the subset of the pooled cohort with adequate follow up to observe dementia at five years the identified ideal cognitive burden score cut point was 0.049 which produced a sensitivity of 0.65 (95% CI 0.64 – 0.67) and specificity of 0.66 (95% CI 0.65 – 0.66) (eTable 2). The cognitive burden score AUC for five-year dementia risk in the pooled cohort was 0.71 (95% CI 0.7 – 0.71). Full cut point analysis results for the pooled cohort, both the individual site cohorts, as well as the younger and older sub-cohorts of the pooled cohort are presented in eTable 2 along with AUCs. In cross sectional sensitivity analysis for the primary aim of assessing association between cognitive burden score and incident dementia diagnosis the odds of dementia at five years per unit increase in cognitive burden score were 3.6 (95% CI 3.4 – 3.7) adjusting for age, sex, race, and comorbidity index (eTable 3). Cognitive burden score remained significantly associated in all sensitivity checks: including multiple encounters per patient (cognitive aOR 3.5, 95% CI 3.4 – 3.6), excluding diagnosis within two years of index admission (cognitive aOR 2.4, 95% CI 2.2 – 2.6), including primary diagnosis category adjustment (cognitive aOR 3.1, 95% CI 3.0 – 3.3), and including primary diagnosis of stroke adjustment (cognitive aOR 3.6, 95% CI 3.4 – 3.8). Finally, logistic regression on five-year dementia risk stratified by quartile of age risk produced consistently positive associations between cognitive burden score and dementia diagnosis risk (eTable 4).

To understand timing of dementia diagnosis and clarify the individual terms contributing to the cognitive symptom score, we next examined whether individual terms from the cognitive score associated with early incident dementia diagnosis (diagnosis within 1 year, versus greater than 1 year) in the pooled cohort. Figure 2 shows the odds of significant (chi square with Bonferroni corrected p < .0007) associations between individual term presence and early (<1 year) versus late (>1 year) diagnosis. Those individual terms with greatest effect on odds of early versus late diagnosis included ‘impulsive’; ‘answer’; ‘forgetful’; ‘cognitive’; and ‘conversation’. Only one term was associated with greater odds of later diagnosis: ‘exercise’.

Finally, as a secondary study we analyzed the cohort of individuals (n=9,872) with a diagnosis of dementia at or before the index hospitalization (i.e., the prevalent dementia cohort), to test whether the cognitive symptom score was associated with all-cause mortality, as a potential marker of disease severity and progression. Characteristics of this cohort, which resulted in 18,232 patient-years of follow-up data and median follow up of 373 days, are summarized in Table 3. In models stratified for age, sex, white race, and Charlson comorbidity score, cognition score quartile was associated with mortality (chi square

(3df)=283; $p < 1e-6$). Median survival ranged from 622 days (95% CI 577–676 days) for the lowest quartile of cognitive symptom scores to 282 days (95% CI 253–310 days) for the highest quartile.

4. DISCUSSION

In this investigation of 535,814 hospitalized individuals without a prevalent diagnosis of dementia and followed for 2,571,793 patient-years, we found that a readily-implemented and interpretable NLP measure of cognitive symptoms, derived in previous work in an independent cohort, was associated with risk for subsequent dementia diagnosis. Individuals in the greatest risk quartile who survived at least five years following hospitalization were 7.4 times more likely to receive a diagnosis of dementia than those in the lowest risk quartile and reached a 1% diagnosis incidence 1,361 days sooner.

Previous studies seeking to predict incident dementia have focused on coded EHR data. In the largest such study, using a Taiwanese database of 387,595 individuals hospitalized for heart failure, two claims-based models performed modestly in predicting dementia.[46] In a smaller-scale German health insurance study of 3,547 individuals, clinician-reported cardiovascular health was associated with incident dementia.[47] Finally, in one preliminary study of 605 medical records, incorporation of NLP-derived features improved discrimination of individuals with dementia from healthy controls.[48]

We further explored the individual terms loading onto this cognitive measure by comparing terms enriched among those diagnosed in the year following discharge rather than subsequently. The associated terms suggest change in features of social interaction (e.g., impulsivity, conversation). Conversely, only one term - 'exercise' - was associated with longer time to diagnosis. Although modest protective effects of exercise on dementia have been reported,[49] whether the present finding reflects a protective factor, a feature leading to later diagnosis, reverse causation, or a proxy for another confounder, merits further study.

Finally, in a second cohort of individuals already diagnosed with dementia at time of hospitalization, we examined whether the cognitive measure was associated with subsequent mortality risk, recognizing that dementia progression as commonly characterized represents a mortality risk factor.[19] Indeed, greater cognitive symptom scores were associated with a greater than 50% reduction in median survival, from 622 days to 282 days. In general, our results comport with other investigations using systematic measures of cognition or measures of frailty or physical functioning[50–52] and adds to the established literature on symptom severity predicting mortality in dementia.[20] While we could not identify a comparable hospital-based study, a recent Taiwanese claims-based study identified predominantly comorbidities and overall utilization to be associated with mortality risk in 37,000 dementia patients.[53] In future work the natural language processing method reported here could be evaluated for disease severity monitoring in cases of recognized dementia allowing for incorporation of patient trajectory through multiple encounters to be studied.

We note several key limitations to be considered. First, while this study suggests the utility of a novel risk factor - cognition symptom burden estimated from NLP - it is entirely possible, even likely, that this prediction could be improved further with the inclusion of additional measures or biomarkers. That is, a more focused set of features might well improve prediction further, and an optimized and targeted machine learning approach to the problem of forecasting future dementia would be an important next step. Our results suggest the utility of incorporating NLP applied to discharge summaries in future accuracy-focused forecasting efforts while at the same time providing an interpretable and accessible symptom score for both risk stratification and severity monitoring. Although NLP gives important insights into non-diagnostic symptoms, the generalizability of models incorporating NLP across health systems remains a challenge; our finding that the same measure exhibits discrimination in two distinct systems is promising in this regard. Although the specific discrimination across sites differs on visual inspection, the two systems are distinct in population and practice and thus variability is to be expected. As this is a feature, not a model, it is possible that this difference could be accounted for through site specific calibration of a predictive model including multiple clinical variables beyond quantified symptoms from the text of a note. Finally, the claims-based outcome is likely imperfect, and may be particularly so for earlier-onset or less common forms of dementia; however, we note that dementia is typically underdiagnosed (which would bias the present study toward the null hypothesis) and prior formal validation has found positive predictive value of dementia codes to be >75% in most health systems.[54,55] Further work is needed on highly scalable computed phenotypes as many existing approaches consume a full medical record and thus are poorly suited to time-to-event analysis.[56,57] The NLP approach reported here could contribute to that effort.

Second, large-scale practice-based evidence of the sort presented here is not a replacement for, or directly comparable to, prospective cohorts with formal longitudinal neuropsychiatric testing and evaluation. The comprehensive representation of two large hospital-based cohorts is a strength of the present work; however, the approach requires use of a coded diagnostic outcome that likely underestimates clinically recognized, but not coded, diagnosis. That is, application of more sensitive measures to define the clinical endpoint would likely identify substantial misclassification, and a prospective investigation applying the methodology we describe would be valuable.

Conversely, the use of previously derived and validated symptom features with known biological correlates is a key strength of this work.[18] That is, the present work is an example of transfer learning, in which a model built in one context is applied in another. The cognition measure used is conceptually simple and readily implemented in a transactional clinical care system or at larger scale for batch population management, making this computed phenotype readily translatable. It is notable that we observe association in two distinct hospital systems, and robust effects across multiple sensitivity analyses. Although limitations in cross-institutional data preclude adjustment for anticholinergic medications, polypharmacy, or use of anticholinesterase inhibitors, as well as other features such as education, this is an opportunity for further research as additional data become available. In addition, the extent to which incorporating data across multiple admissions may improve prediction remains to be explored.

In sum, this study of more than 500,000 individuals drawn from two academic medical centers demonstrates the potential utility of a cognitive score derived from NLP of discharge summaries in predicting incident dementia, as well as mortality risk among individuals with dementia. Prior to formal diagnosis of dementia, symptoms documented in discharge summaries may facilitate identification of high-risk individuals in whom further evaluation or application of biomarkers may have greatest yield. More generally, our results suggest the extent to which scalable efforts to derive concepts from discharge notes may be used to supplement administrative data in clinical settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Funding

This work was supported by the National Institute of Mental Health (grant number 1R01MH106577), the National Institute of Aging (Supplement to R01MH104488), and the Brain and Behavior Research Foundation. The sponsors had no role in study design, writing of the report, or data collection, analysis, or interpretation.

Declaration of Interest

THM receives research funding from the Stanley Center at the Broad Institute, the Brain and Behavior Research Foundation, National Institute of Aging, and Telefonica Alfa. RHP holds equity in Psy Therapeutics and Outermost Therapeutics; serves on the scientific advisory boards of Genomind and Takeda; consults to RID Ventures; and is an Associate Editor at *JAMA Network Open*. RHP receives research funding from NIMH, NHLBI, NHGRI, and Telefonica Alfa. The other authors have no disclosures to report.

REFERENCES

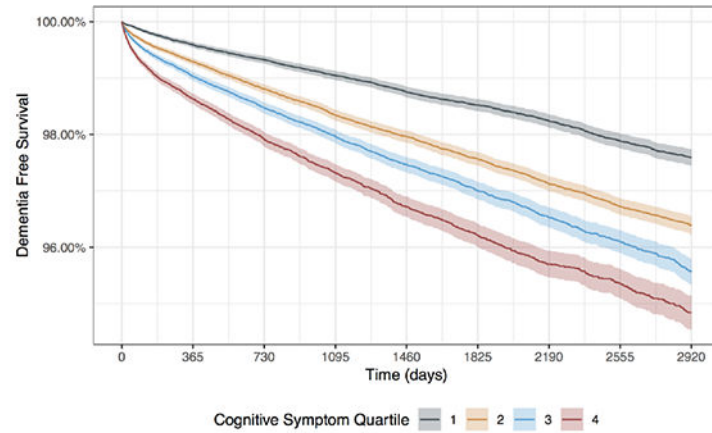
- [1]. Birks JS, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst Rev* 2018;6:CD001190. doi:10.1002/14651858.CD001190.pub3.
- [2]. Blanco-Silvente L, Capellà D, Garre-Olmo J, Vilalta-Franch J, Castells X. Predictors of discontinuation, efficacy, and safety of memantine treatment for Alzheimer's disease: meta-analysis and meta-regression of 18 randomized clinical trials involving 5004 patients. *BMC Geriatr* 2018;18:168. doi:10.1186/s12877-018-0857-5. [PubMed: 30041625]
- [3]. Tricco AC, Ashoor HM, Soobiah C, Rios P, Veroniki AA, Hamid JS, et al. Comparative Effectiveness and Safety of Cognitive Enhancers for Treating Alzheimer's Disease: Systematic Review and Network Metaanalysis. *J Am Geriatr Soc* 2018;66:170–8. doi:10.1111/jgs.15069. [PubMed: 29131306]
- [4]. Reiman EM, Langbaum JB, Tariot PN, Lopera F, Bateman RJ, Morris JC, et al. CAP—advancing the evaluation of preclinical Alzheimer disease treatments. *Nat Rev Neurol* 2016;12:56–61. doi:10.1038/nrneurol.2015.177. [PubMed: 26416539]
- [5]. Ossenkoppele R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BNM, et al. Prevalence of Amyloid PET Positivity in Dementia Syndromes: A Meta-analysis. *JAMA* 2015;313:1939–49. doi:10.1001/jama.2015.4669. [PubMed: 25988463]
- [6]. Frisoni GB, Bocchetta M, Chetelat G, Rabinovici GD, de Leon MJ, Kaye J, et al. Imaging markers for Alzheimer disease: Which vs how. *Neurology* 2013;81:487–500. doi:10.1212/WNL.0b013e31829d86e8. [PubMed: 23897875]
- [7]. Watson JL, Ryan L, Silverberg N, Cahan V, Bernard MA. Obstacles And Opportunities In Alzheimer's Clinical Trial Recruitment. *Health Aff* 2014;33:574–9. doi:10.1377/hlthaff.2013.1314.

- [8]. Becker RE, Greig NH. Alzheimer's disease drug development in 2008 and beyond: problems and opportunities. *Curr Alzheimer Res* 2008;5:346–57. [PubMed: 18690832]
- [9]. Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY. Cognitive Tests to Detect Dementia: A Systematic Review and Meta-analysis. *JAMA Intern Med* 2015;175:1450–8. doi:10.1001/jamainternmed.2015.2152. [PubMed: 26052687]
- [10]. Ford E, Greenslade N, Paudyal P, Bremner S, Smith HE, Banerjee S, et al. Predicting dementia from primary care records: A systematic review and meta-analysis. *PLoS One* 2018;13:e0194735. doi:10.1371/journal.pone.0194735. [PubMed: 29596471]
- [11]. Uzuner O, Stubbs A, Filannino M. A natural language processing challenge for clinical records: Research Domains Criteria (RDoC) for psychiatry. *J Biomed Inform* 2017;75S:S1–S3. doi:10.1016/j.jbi.2017.10.005. [PubMed: 29042245]
- [12]. McCoy TH Jr, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. *JAMA Psychiatry* 2016;73:1064–71. doi:10.1001/jamapsychiatry.2016.2172. [PubMed: 27626235]
- [13]. McCoy TH Jr, Pellegrini AM, Perlis RH. Research Domain Criteria scores estimated through natural language processing are associated with risk for suicide and accidental death. *Depress Anxiety* 2019. [Epub ahead of print]. doi: 10.1002/da.22882.
- [14]. McCoy TH Jr, Castro VM, Cagan A, Roberson AM, Kohane IS, Perlis RH. Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study. *PLoS One* 2015;10:e0136341. doi:10.1371/journal.pone.0136341. [PubMed: 26302085]
- [15]. Wasfy JH, Strom JB, Waldo SW, O'Brien C, Wimmer NJ, Zai AH, et al. Clinical Preventability of 30-Day Readmission After Percutaneous Coronary Intervention. *J Am Heart Assoc* 2014;3:e001290. doi:10.1161/JAHA.114.001290. [PubMed: 25261531]
- [16]. McCoy TH Jr, Castro VM, Rosenfield HR, Cagan A, Kohane IS, Perlis RH. A clinical perspective on the relevance of research domain criteria in electronic health records. *Am J Psychiatry* 2015;172:316–20. doi:10.1176/appi.ajp.2014.14091177. [PubMed: 25827030]
- [17]. McCoy TH Jr, Yu S, Hart KL, Castro VM, Brown HE, Rosenquist JN, et al. High Throughput Phenotyping for Dimensional Psychopathology in Electronic Health Records. *Biol Psychiatry* 2018;83:997–1004. doi:10.1016/j.biopsych.2018.01.011. [PubMed: 29496195]
- [18]. McCoy TH Jr, Castro VM, Hart KL, Pellegrini AM, Yu S, Cai T, et al. Genome-wide Association Study of Dimensional Psychopathology Using Electronic Health Records. *Biol Psychiatry* 2018;83:1005–11. doi:10.1016/j.biopsych.2017.12.004. [PubMed: 29496196]
- [19]. Mitchell SL, Miller SC, Teno JM, Kiely DK, Davis RB, Shaffer ML. Prediction of 6-month survival of nursing home residents with advanced dementia using ADEPT vs hospice eligibility guidelines. *JAMA* 2010;304(17):1929–1935. doi:10.1001/jama.2010.1572. [PubMed: 21045099]
- [20]. Connors MH, Ames D, Boundy K, Clarnette R, Kurrle S, Mander A, Ward J, Woodward M, Brodaty H. Predictors of Mortality in Dementia: The PRIME Study. *J Alzheimers Dis* 2016;52(3):967–974. doi:10.3233/JAD-150946. [PubMed: 27079702]
- [21]. Kuperman G, Bates DW. Standardized coding of the medical problem list. *J Am Med Inform Assoc JAMIA* 1994;1:414–5. [PubMed: 7850566]
- [22]. Wright A, McCoy AB, Hickman T-TT, Hilaire DS, Borbolla D, Bowes WA, et al. Problem list completeness in electronic health records: A multi-site study and assessment of success factors. *Int J Med Inf* 2015;84:784–90. doi:10.1016/j.ijmedinf.2015.06.011.
- [23]. Kovalchuk Y, Stewart R, Broadbent M, Hubbard TJP, Dobson RJB. Analysis of diagnoses extracted from electronic health records in a large mental health case register. *PLoS One* 2017;12:e0171526. doi:10.1371/journal.pone.0171526. [PubMed: 28207753]
- [24]. Wilkinson T, Ly A, Schnier C, Rannikmäe K, Bush K, Brayne C, et al. Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimers Dement* 2018;14:1038–51. doi:10.1016/j.jalz.2018.02.016. [PubMed: 29621480]
- [25]. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. *AMIA Annu Symp Proc* 2007:548–52. [PubMed: 18693896]

- [26]. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc JAMIA* 2010;17:124–30. doi:10.1136/jamia.2009.000893. [PubMed: 20190053]
- [27]. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010;167(7):748–51. doi:10.1176/appi.ajp.2010.09091379. [PubMed: 20595427]
- [28]. Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. *Science* 2015;348(6234):499–500. doi:10.1126/science.aab2358. [PubMed: 25931539]
- [29]. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94:496–509.
- [30]. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med* 2017;36(27):4391–4400. doi:10.1002/sim.7501. [PubMed: 28913837]
- [31]. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016;133(6):601–9. doi:10.1161/CIRCULATIONAHA.115.017719. [PubMed: 26858290]
- [32]. Kawaguchi ES, Shen JI, Li G, Suchard MA. A Fast and Scalable Implementation Method for Competing Risks Data with the R Package fastcmprsk. *Journal of Statistical Software* 2019;10(2):1–21. doi:10.18637/jss.v000.i00.
- [33]. McCoy TH Jr, Wiste AK, Doyle AE, Pellegrini AM, Perlis RH. Association between child psychiatric emergency room outcomes and dimensions of psychopathology. *Gen Hosp Psychiatry* 2019;59:1–6. doi:10.1016/j.genhosppsych.2019.04.009. [PubMed: 31034963]
- [34]. McCoy TH Jr, Pellegrini AM, Perlis RH. Research Domain Criteria scores estimated through natural language processing are associated with risk for suicide and accidental death. *Depress Anxiety* 2019;36(5):392–399. doi:10.1002/da.22882. [PubMed: 30710497]
- [35]. McCoy TH Jr, Yu S, Hart KL, Castro VM, Brown HE, Rosenquist JN, Doyle AE, Vuijk PJ, Cai T, Perlis RH. High Throughput Phenotyping for Dimensional Psychopathology in Electronic Health Records. *Biol Psychiatry* 2018;83(12):997–1004. doi:10.1016/j.biopsych.2018.01.011. [PubMed: 29496195]
- [36]. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* 2010;143(3):331–6. doi:10.1016/j.otohns.2010.05.007. [PubMed: 20723767]
- [37]. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3(2):143–52. [PubMed: 6463451]
- [38]. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247(18):2543–6. [PubMed: 7069920]
- [39]. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;30(10):1105–17. doi:10.1002/sim.4154. [PubMed: 21484848]
- [40]. Bamber D The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology.* 1975;12:387–415.
- [41]. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29–36. [PubMed: 7063747]
- [42]. Unal I Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Comput Math Methods Med* 2017;3762651. doi:10.1155/2017/3762651. [PubMed: 28642804]
- [43]. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3(1):32–5. [PubMed: 15405679]
- [44]. Miha V, Tomaz C. ROC Curve, Lift Chart and Calibration Plot. *Metodološki zvezki* 2006;3(1)89–108.
- [45]. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
- [46]. Hu W-S, Lin C-L. Comparison of CHA2DS2-VASc and AHEAD scores for the prediction of incident dementia in patients hospitalized for heart failure: a nationwide cohort study. *Intern Emerg Med* 2018 [Epub ahead of print]. doi:10.1007/s11739-018-1961-4.

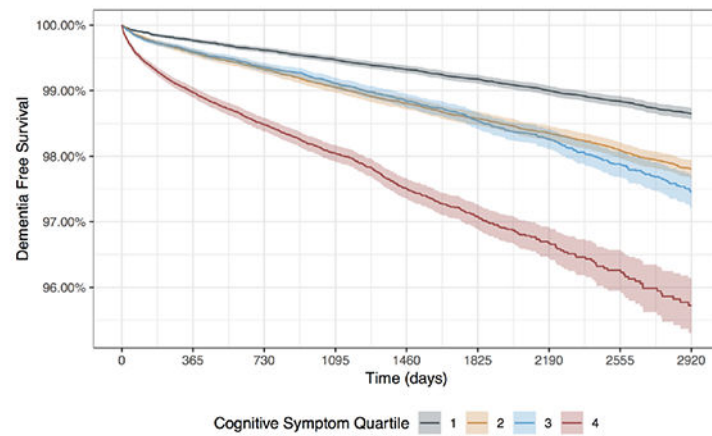
- [47]. Hessler JB, Ander K-H, Brönnner M, Etgen T, Förstl H, Poppert H, et al. Predicting dementia in primary care patients with a cardiovascular health metric: a prospective population-based study. *BMC Neurol* 2016;16:116. doi:10.1186/s12883-016-0646-8. [PubMed: 27459854]
- [48]. Moreira LB, Namen AA. A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Comput Methods Programs Biomed* 2018;165:139–49. doi:10.1016/j.cmpb.2018.08.016. [PubMed: 30337069]
- [49]. Lautenschlager NT, Cox KL, Flicker L, Foster JK, van Bockxmeer FM, Xiao J, et al. Effect of Physical Activity on Cognitive Function in Older Adults at Risk for Alzheimer Disease: A Randomized Trial. *JAMA* 2008;300:1027–37. doi:10.1001/jama.300.9.1027. [PubMed: 18768414]
- [50]. Forti P, Maioli F, Lega MV, Montanari L, Coraini F, Zoli M. Combination of the Clock Drawing Test with the Physical Phenotype of Frailty for the Prediction of Mortality and Other Adverse Outcomes in Older Community Dwellers without Dementia. *Gerontology* 2014;60:204–11. doi:10.1159/000356701. [PubMed: 24356341]
- [51]. Verdan C, Casarsa D, Perroux MR, Santos M, Alves de Souza J, Nascimento O, et al. Lower mortality rate in people with dementia is associated with better cognitive and functional performance in an outpatient cohort. *Arq Neuropsiquiatr* 2014;72:278–82. [PubMed: 24760091]
- [52]. Paradise M, Walker Z, Cooper C, Blizard R, Regan C, Katona C, et al. Prediction of survival in Alzheimer's disease—the LASER-AD longitudinal study. *Int J Geriatr Psychiatry* 2009;24:739–47. doi:10.1002/gps.2190. [PubMed: 19189277]
- [53]. Lee K-C, Hsu W-H, Chou P-H, Yiin J-J, Muo C-H, Lin Y-P. Estimating the survival of elderly patients diagnosed with dementia in Taiwan: A longitudinal study. *PLoS One* 2018;13:e0178997. doi:10.1371/journal.pone.0178997. [PubMed: 30044781]
- [54]. 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. [PubMed: 29508259]
- [55]. Wilkinson T, Ly A, Schnier C, Rannikmae K, Bush K, Brayne C, Quinn TJ, Sudlow CLM, UK Biobank Neurodegenerative Outcomes Group and Dementias Platform UK. Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimers Dement* 2018;14(8):1038–1051. doi:10.1016/j.jalz.2018.02.016. [PubMed: 29621480]
- [56]. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013;31(12):1102–10. doi:10.1038/nbt.2749. [PubMed: 24270849]
- [57]. Yu S, Ma Y, Gronsbell J, Cai T, Ananthakrishnan AN, Gainer VS, Churchill SE, Szolovits P, Murphy SN, Kohane IS, Liao KP, Cai T. Enabling phenotypic big data with PheNorm. *J Am Med Inform Assoc* 2018;25(1):54–60. doi:10.1093/jamia/ocx111. [PubMed: 29126253]

A (Site 1)



Time (days)	Risk								
	0	365	730	1095	1460	1825	2190	2555	2920
1 st Quartile	57295	54246	52157	48048	44096	40455	36798	32593	28464
2 nd Quartile	74537	68588	64981	59027	53707	48467	42761	35176	27913
3 rd Quartile	69892	58556	48956	40716	33130	26122	19699	15497	11823
4 th Quartile	66131	45458	32985	26653	20916	15847	11461	8972	6882

B (Site 2)



Time (days)	Risk								
	0	365	730	1095	1460	1825	2190	2555	2920
1 st Quartile	76659	72842	71113	69833	68718	67733	66752	60041	52216
2 nd Quartile	59416	53975	51521	49502	47622	45675	42955	35970	28428
3 rd Quartile	64062	50455	40308	33655	26913	19036	10819	7956	5947
4 th Quartile	67822	49536	38153	28590	19386	11031	4602	2976	1939

Figure 1 A-B: Kaplan-Meier plots of time to dementia outcome in following index hospitalization over 8 years of follow-up by quartile of cognitive symptom burden score (color) for each hospital system (panel).

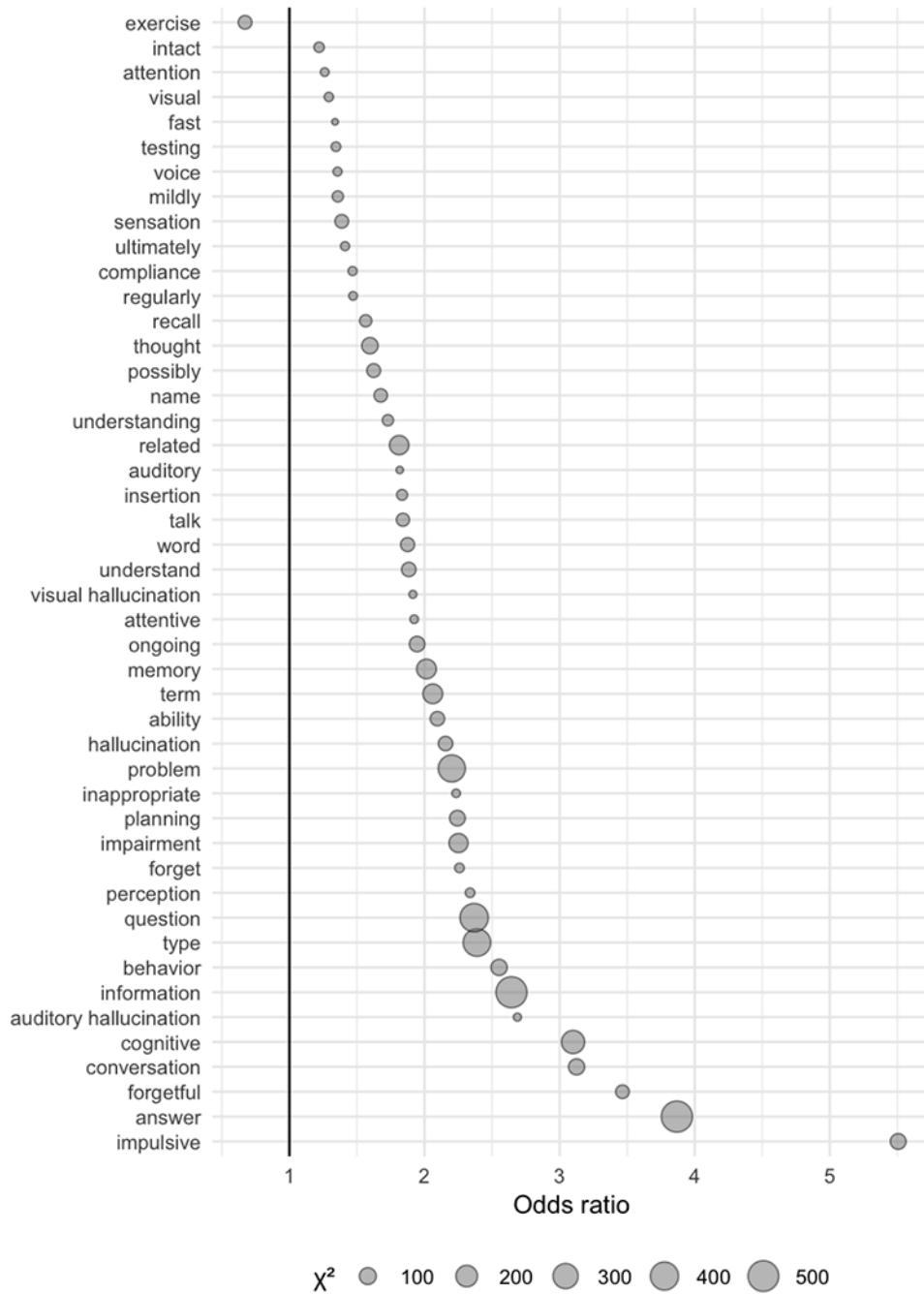


Figure 2: Terms from the cognitive symptom score where the presence or absence of which is individually significantly associated with early (<1 year) versus late (>1 year) incident dementia diagnosis in the pooled cohort.

Table 1:

Characteristics of incident dementia cohorts studied for association between symptom burden and eventual diagnosis of dementia at each hospital site and pooled.

	Site 1	Site 2	Pooled
<i>N</i>	267855	267959	535814
Sex (male, %)	124410 (46.4)	94767 (35.4)	219177 (40.9)
Race (white, %)	212670 (79.4)	196437 (73.3)	409107 (76.4)
Age (years, mean (sd))	54.88 (18.93)	51.81 (18.75)	53.34 (18.90)
Charlson Comorbidity Index (mean (sd))	4.05 (4.29)	3.22 (3.74)	3.63 (4.05)
Dementia Outcomes (%)	6516 (2.43)	3902 (1.46)	10418 (1.94)
Primary diagnosis of stroke (%)	11958 (4.5)	6723 (2.5)	18681 (3.5)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2a:

Fine-Gray subdistribution hazards model (death as competing risk for dementia diagnosis), adjusted for age, sex, race, and cognitive score at Site 1.

	aHR	95% CI	p-value
Age (years)	1.064	(1.062, 1.065)	<0.0001
Sex (ref = female)	0.912	(0.874, 0.952)	<0.0001
Race (ref = white)	1.081	(1.010, 1.157)	0.024
Charlson (log)	5.210	(4.759, 5.703)	<0.0001
Cognitive Symptom	1.626	(1.535, 1.722)	<0.0001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2b:

Fine-Gray subdistribution hazards model (death as competing risk for dementia diagnosis), adjusted for age, sex, race, and cognitive score at Site 2.

	aHR	95% CI	p-value
Age (years)	1.064	(1.062, 1.067)	<0.0001
Sex (ref = female)	0.911	(0.860, 0.965)	0.0016
Race (ref = white)	1.603	(1.488, 1.727)	<0.0001
Charlson (log)	4.674	(4.123, 5.297)	<0.0001
Cognitive Symptom	1.374	(1.290, 1.463)	<0.0001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Characteristics of the prevalent dementia cohort studied for association between symptom burden in the context of recognized dementia and all-cause mortality.

	Pooled
<i>N</i>	9872
Sex (male, %)	4393 (44.5)
Race (white, %)	8157 (82.6)
Age (years, mean (sd))	82.55 (9.41)
Charlson Comorbidity Index (mean (sd))	9.78 (5.07)
Mortality Outcomes (%)	8377 (84.86)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript