OXFORD

Research Practice

Validation of Claims Algorithms to Identify Alzheimer's Disease and Related Dementias

Ellen P. McCarthy, PhD, MPH,[1,2] **Chiang-Hua Chang, PhD, MS,**[3,4] **Nicholas Tilton, PhD,**[3] **Mohammed U. Kabeto, MS,**[3,5] **Kenneth M. Langa, MD, PhD,**[3,5] **and Julie P. W. Bynum, MD, MPH**[3,4,*]

[1]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts, USA. [2]Division of Gerontology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA. [4]Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan, USA. [5]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA.

*Address correspondence to: Julie P. W. Bynum, MD, MPH, Department of Internal Medicine, University of Michigan, 2800 Plymouth Road, NCRC-16 Room 446E, Ann Arbor, MI 48109-2800, USA. E-mail: Bynumju@med.umich.edu

Abstract

Background: Using billing data generated through health care delivery to identify individuals with dementia has become important in research. To inform tradeoffs between approaches, we tested the validity of different Medicare claims-based algorithms.
Methods: We included 5 784 Medicare-enrolled, Health and Retirement Study participants aged older than 65 years in 2012 clinically assessed for cognitive status over multiple waves and determined performance characteristics of different claims-based algorithms.
Results: Positive predictive value (PPV) of claims ranged from 53.8% to 70.3% and was highest using a revised algorithm and 1 year of observation. The tradeoff of greater PPV was lower sensitivity; sensitivity could be maximized using 3 years of observation. All algorithms had low sensitivity (31.3%–56.8%) and high specificity (92.3%–98.0%). Algorithm test performance varied by participant characteristics, including age and race.
Conclusion: Revised algorithms for dementia diagnosis using Medicare administrative data have reasonable accuracy for research purposes, but investigators should be cognizant of the tradeoffs in accuracy among the approaches they consider.

Keywords: Accuracy, Algorithm, Dementia, Diagnosis, Medicare

Background

Alzheimer's disease and related dementias (ADRDs, hereafter referred to as dementia) are debilitating conditions characterized by a decline in memory and other cognitive domains leading to progressive loss of independence. There are 6.2 million Americans living with dementia in 2021, nearly two thirds are 75 years and older and 70% are community-dwelling (1). With the baby boomer generation turning age 75 in 2021, this number will expand rapidly to 7.1 million by 2025 and 13.8 million by 2050 (1,2). Given these projections, it is critically important for clinical improvement, health care research, and reimbursement to be able to identify people living with dementia accurately using billing data generated through the delivery of care (3,4). These data are also used to identify people living

with dementia for pragmatic clinical trials, population management, and to update population prevalence estimates because they are inclusive of all older adults who receive care through Medicare entitlement, including groups who may not typically volunteer for studies, such as racial and ethnic minorities (5,6).

The Center for Medicare and Medicaid Services' Chronic Conditions Warehouse (CCW) created an algorithm to support the use of Medicare claims data to identify people with dementia based on validation studies of billing practices in the 1990s (7–9). With changes in clinical practice and greater attention on dementia, the test performance of claims may have changed, especially now that more diagnoses can be documented on bills. The CCW algorithm requires only one claim to address potential underdiagnosis, which

may lead to lower specificity, and it also does not include Lewy Body dementia. To increase sensitivity, the CCW algorithm uses 3 years of consecutive Medicare claims data (7,9). But this longer observation period is often not practical for uses such as enrollment in pragmatic trials or in policy and health services research.

In this context, we tested the performance of new algorithms to identify individuals with dementia that address limitations of the CCW algorithm and inform the tradeoffs among different dementia algorithms using the Health and Retirement Study (HRS) as the reference standard. We further assessed the validity of 1- and 3-year observation periods and stratified by population subgroups (ie, age, gender, race/ethnicity, and respondent type). This evaluation focused on the time period when ICD-9 codes were in use to reduce the number of factors that may alter validity, but will serve as a point of comparison for future evaluations of the International Classification of Diseases, 10th Revision (ICD-10) coding system.

Method

Data Source

We used data from the HRS linked to Medicare enrollment and claims data. The HRS is a nationally representative, longitudinal study of adults aged 51 and older interviewed biennially since 1992 with waves of new participants enrolled periodically to maintain a steady-state cohort. To minimize loss to follow-up, proxy interviews are conducted, usually with a spouse or family member, if a respondent is unable or unwilling to complete a self-interview. We obtained Medicare data from 2010 to 2014 (ie, the last full year of ICD-9-CM coding in the United States) for beneficiaries aged 65 and older enrolled in the Medicare fee-for-service (FFS) program.

Informed consent was obtained from all participants upon study enrollment and upon Medicare eligibility for linkage to Medicare data; 88% of participants consented to Medicare linkage (10). The Institutional Review Board at the University of Michigan approved this study.

Study Participants

We studied participants of the 2012 survey wave and constructed 2 observation periods to look for ADRD diagnoses in Medicare claims. The 1-year window comprised 5 784 participants who were aged 65.5 years or older and continuously enrolled in Medicare FFS Part A and Part B for 6 months before and 6 months following their HRS interview or until death. The 3-year window comprised 5 315 participants with the same eligibility, but for 18 months before and 18 months following their HRS interview or until death.

We obtained sociodemographic and health characteristics of the cohort in 2012, including age, sex, self-reported race/ethnicity, educational attainment, residence in the community or nursing home/other facilities, urban/rural location, and respondent type (self or proxy). Health factors included limitations in activities of daily living (ADLs) and instrumental activities of daily living (IADLs), self or proxy-reported health status, and CCW chronic conditions (claims-based flags from the chronic conditions segment of the Master Beneficiary Summary File). We also report death within 6 months and 1 year of the HRS interview date.

HRS Cognitive Status—Reference Standard

We used the HRS cognitive status as the reference standard using validated methods developed by Langa, Kabeto, and Weir (LKW method) (11,12). This method uses validated cognitive and functional assessments of participants to label individuals as having dementia, cognitive impairment not dementia (CIND), or normal cognitive function. For self-respondents, the LKW method uses measures of immediate and delayed word recall, serial-7 subtraction, and backward counting from 20 using the modified Telephone Interview for Cognitive Status. These measures are used to derive a 27-point cognitive scale classifying self-respondents as having dementia (0–6 points), CIND (7–11 points), or normal (12–27), with cutpoints validated against the Aging, Demographics, and Memory Study (ADAMS), an HRS substudy that included in-person neuropsychological and clinical assessments with expert adjudication to determine dementia diagnosis (13). For participants with proxy respondents, the LKW method uses INFORMATION ON whether cognitive limitation was the reason for a proxy interview, and the proxy's assessment of the participant's memory and difficulties with IADLs. The proxy cognition score is an 11-point scale classifying participants as having dementia (6–11 points), CIND (3–5 points), or normal (0–2) (11,12).

HRS cognitive status for each participant was assessed across 4 survey waves (2006–2012). To reduce measurement error related to cognitive test performance, a participant identified as having dementia with continued evidence of dementia or CIND in any wave was considered a dementia case in 2012 (14–16). CIND was assessed in 2012 alone because this reflects a more transient state, whereby some people get better, some deteriorate, and others stay the same.

CCW ADRD Algorithm

The CCW ADRD flag is available to researchers on the chronic condition segment of the Medicare Beneficiary Annual Summary File. This flag is based on an algorithm developed by Tayor et al. USING CLAIMS FROM the early 1990s (7) and validated in ADAMS using Medicare claims through 2005 (8). The CCW ADRD algorithm uses a 3-year observation period to flag records of beneficiaries who had at least one eligible Medicare claim with an ICD-9-CM diagnosis code listed in Table 1. Eligible claims include any hospital, skilled nursing facility, hospital outpatient, and home health episode as well as any Carrier file service (professional services, labs, and tests) except for durable medical equipment and ambulance services, which are excluded. Hospice claims are not included (9).

Bynum ADRD Algorithms

We evaluated whether modifications to the CCW algorithm improved classification of ADRD status by (a) shortening the observation period from 3 years to 1 year; (b) adding diagnosis codes for dementia with Lewy Bodies (331.82), other cerebral degeneration (331.89), and other nonspecified senile psychosis (290.8) based on discussion with experts in the field; and (c) modifying the claims input files by adding hospice claims and including only encounters in the hospital outpatient file (HOF) by underserved populations who receive care from Federally Qualified Health Centers, Rural Health Centers, and Critical Access Hospitals under payment option II.

We then constructed 2 new algorithms designed to address the potential for low specificity of the CCW algorithm (Table 1). In both algorithms, individuals are flagged with dementia if there is at least one qualifying claim for hospital inpatient, skilled nursing facility, home health care, or hospice service. In the first algorithm (Bynum-EM), beneficiaries could additionally be flagged if they had at least one claim for a face-to-face patient visit by a physician or other clinician determined by Berenson-Eggers Type of Service codes for "evaluation and management" (EM) services in the Carrier file or a qualifying visit in the HOF file. The second

Table 1. ADRD Algorithm Specifications

	CCW	Bynum-EM	Bynum-Standard
Observation Period	3 Years	1 Year and 3 Years	1 Year and 3 Years
ICD-9-CM Diagnosis Codes	331.0, 331.11, 331.19, 331.2, 331.7, 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, 294.0, 294.10, 294.11, 294.20, 294.21, 294.8, 797	331.0, 331.11, 331.19, 331.2, 331.7, **331.82, 331.89,** 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, **290.8,** 294.0, 294.10, 294.11, 294.20, 294.21, 797	331.0, 331.11, 331.19, 331.2, 331.7, **331.82, 331.89,** 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, **290.8,** 294.0, 294.10, 294.11, 294.20, 294.21, 797
Claims Files and Qualifying Claims			
MEDPAR	Any inpatient or SNF claim	Any inpatient or SNF claim	Any inpatient or SNF claim
Home Health Agency	Any claim*	Any claim	Any claim
Hospice	Any claim*	Any claim	Any claim
HOF for outpatient medical services	Any claim*	Includes only claims from Rural Health Clinics, Federally Qualified Health Centers, and Critical Access Hospitals—Payment Option II	Includes only claims from Rural Health Clinics, Federally Qualified Health Centers, and Critical Access Hospitals—Payment Option II
Carrier (Provider) File for services from physicians and other health care providers	Any claim*, *excluding claims* with BETOS codes of D1A, D1B, D1C, D1D, D1E, D1F, D1G (for durable medical equipment), or O1A (for ambulance services)	Any claim for evaluation and management (E&M) by a provider *Includes only claims* BETOS "M" codes: M1A, M1B, M2A, M2B, M2C, M3, M4A, M4B, M5A, M5B, M5C, or M6	Any claim *This algorithm requires two or more qualifying Carrier or HOF claims at least 7 days apart.*

Note: ADRD = Alzheimer's disease and related dementias; CCW = Chronic Conditions Warehouse; EM = Evaluation and Management; ICD-9-CM = International Classification of Diseases, 9th Edition, Clinical Modification; MEDPAR = Medical Provider and Analysis File; SNF = skilled nursing facility; HOF = hospital outpatient file; BETOS = Berenson-Eggers Type of Service. The bolded numbers indicate diagnostic codes that were added to the Bynum algorithms.
*NCH Claim Type code includes home health agency claim = 10; hospital outpatient claim = 40; hospice claim = 50; Carrier claim = 71 or 72.

algorithm (Bynum-Standard) mimics other comorbidity algorithms by requiring 2 claims for any type of service in the Carrier file or qualifying HOF encounters by underserved populations described above that were at least 7 days apart to account for potential misclassification resulting from "rule out" diagnoses (15–17).

Statistical Analysis

We described the characteristics of the HRS 2012 FFS cohort by cognitive status. We then conducted a head-to-head comparison of the performance of the claims-based algorithms (CCW, Bynum-EM, Bynum-Standard in 3- and 1-year claims observation periods) using the HRS-designated dementia status as the reference standard. We report sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and percent agreement. We also report the Youden Index, a summary measure that combines sensitivity and specificity, which gives equal weight to false positives and false negatives and reflects the proportion of total misclassified results (18). We selected the algorithm with the best PPV and assessed its performance by population subgroup (ie, by age, gender, race/ethnicity, and respondent type). We then compared the claims-identified dementia population prevalence and characteristics to the population with dementia identified by the HRS LKW method.

In secondary analyses, we assessed the performance of (a) the best performing claims-based algorithm identified above against HRS "Dementia or CIND" as the reference standard as opposed to Dementia only and (b) the Mild Cognitive Impairment diagnosis (ICD-9 code 331.83) against the HRS cognitive status of "CIND."

All analyses were performed using SAS 9.4 survey procedures (SAS Institute, Cary, NC) to account for the complex sampling design of HRS. Sampling weights were constructed for this analysis so that the sample of age-eligible Medicare enrollees with claims linkage reflect national estimates of Medicare beneficiaries aged 65 years or older when the weights are applied. Overall, 98% of HRS respondents aged 65 years or older are enrolled in Medicare, and 93% of those individuals are successfully linked to their Medicare records. However, HRS linkage rates vary by age with linkage rates exceeding 97% for respondents aged 85 and older, but are only 86% for those aged 65–69. Linkage rates are also about 5% lower among Black respondents but shown to be generally unrelated to cognitive status (19). New sampling weights were constructed to take into account the probabilities of selection and to adjust for the characteristics related to differential nonlinkage in HRS Medicare data. All results are weighted to reflect national estimates; unweighted sample sizes are presented when appropriate.

Results

Table 2 presents the characteristics of the eligible FFS HRS 2012 cohort by cognitive status for the 1-year cohort, which included 5 784 participants representing an estimated 25.5 million Medicare beneficiaries nationwide. The prevalence of dementia was 12.9% in this cohort. Compared to those with normal cognitive status, beneficiaries with dementia and CIND were more likely to be older, of non-White race, have lower educational attainment, and have chronic conditions. Beneficiaries with dementia more often resided in a nursing home or other health care facility, had impairments on 3 or more ADLs and IADLs, reported fair or poor health, and were represented by a proxy. Overall, 16.4% of beneficiaries with dementia had died within 1 year of their interview compared to 2.5% with normal cognition. Characteristics of the 3-year cohort were similar (Supplementary Table 1).

Table 2. Characteristics of HRS Participants in the 2012 Wave Enrolled in FFS Medicare for 1-Year Claims Observation Period

Characteristics	HRS Cognitive Status				
	Dementia (%)	CIND (%)	Normal (%)	Total	p*
HRS sample size	927	1,075	3,782	5,784	
Weighted N in millions (%)	3.2M (12.9)	4.1M (16.9)	17.2M (70.2)	24.5M (100)	
Demographics					
Age in years, mean (SD)	83.1 (8.0)	78.9 (8.0)	74.8 (6.8)	76.5 (7.8)	.001
Female	64.1	56.8	57.7	58.4	.009
Race/Ethnicity†					<.0001
Non-Hispanic White	70.3	79.5	90.2	85.8	
Non-Hispanic Black	16.4	10.7	4.4	7.0	
Hispanic	10.6	7.5	3.7	5.3	
Education					<.0001
Less than high school	49.0	32.1	11.3	19.7	
High school and equivalent	29.3	36.1	35.5	34.8	
Some college	12.5	19.2	23.6	21.4	
College graduate and above	9.2	12.6	29.6	24.1	
Residential setting					<.0001
Community	78.9	97.2	99.5	96.5	
Nursing home or other health care facility	21.1	2.8	0.5	3.5	
Urban/rural‡					.68
Urban	72.2	72.6	73.6	73.3	
Rural	27.8	27.4	26.4	26.7	
Interview respondent					<.0001
Self	61.0	96.2	98.8	93.5	
Proxy	39.0	3.8	1.2	6.5	
Functional status					
ADL‡ limitations					<.0001
0	32.2	59.1	78.9	69.5	
1 or 2	25.1	25.5	16.2	18.9	
3 or more	42.7	15.4	4.9	11.6	
IADL‡ limitations					<.0001
0	34.6	70.9	88.7	78.7	
1 or 2	18.5	23.0	9.9	13.2	
3 or more	46.9	6.1	1.4	8.1	
Self-reported health status‡					<.0001
Excellent, very good, or good	40.7	60.3	77.2	69.6	
Fair or poor	59.3	39.7	22.8	30.4	
CCW chronic condition flags					
Acute MI/ischemic heart disease	69.0	62.3	45.1	51.1	<.0001
Asthma/COPD	42.3	37.6	26.3	30.3	<.0001
Atrial fibrillation	23.9	20.8	13.7	16.2	<.0001
Chronic kidney disease	38.5	30.9	20.2	24.4	<.0001
Diabetes	47.6	45.4	31.7	36.1	<.0001
Heart failure	50.5	36.8	21.1	27.6	<.0001
Hip or pelvic fracture/osteoporosis	39.3	27.8	21.0	24.5	<.0001
Hypertension	91.6	87.5	77.0	80.7	<.0001
Stroke/TIA	32.4	22.8	11.8	16.3	<.0001

Table 2. Continued

Characteristics	HRS Cognitive Status				
	Dementia (%)	CIND (%)	Normal (%)	Total	p*
Died					
Within 6 months after interview	7.7	2.6	1.1	2.2	<.0001
Within 1 year after interview	16.4	6.1	2.5	4.9	<.0001
Cognition scales, mean (SD)					
Self-responders (n = 5 302)					
TICS-m score§	6.1 (3.4)	9.5 (1.3)	16.4 (2.9)	14.3 (4.4)	<.0001
Proxy-responders (n = 482)					
Proxy cognition score‖	9.2 (2.0)	4.1 (0.8)	1.5 (0.6)	7.7 (0.9)	<.0001

Notes: HRS = Health and Retirement Study; FFS = fee for service; CIND = cognitive impairment not dementia; SD = standard deviation; ADL = activities of daily living; IADL = instrumental activities of daily living; CCW = Chronic Condition Data Warehouse; MI = myocardial infarction; COPD = chronic obstructive pulmonary disease; TIA = transient ischemic attack; TICS-m = modified Telephone Interview for Cognitive Status. The reported percentages are weighted using the authors' sampling weights, which were constructed to account for characteristics of respondents who did not have linked Medicare data. The weighted *N* reflects the population estimate in millions (eg, the prevalence of dementia is 12.9%, representing an estimated 3.2M people and prevalence of CIND is 16.9%, representing 4.1M people nationally. All percentages reflect national population estimates.

*p value tests for the overall association between HRS dementia status and characteristics.

†Percentages do not add to 100% because other race/ethnicity is not presented due to small sample sizes.

‡Missing data: Urban/rural residence (*n* = 99); ADL (*n* = 2); IADL (*n* = 6); self-reported health status (*n* = 5).

§TICS-m score (range, 0–27): 0–6 is consistent with dementia, 7–11 is cognitive impairment without dementia, and 12–27 is consistent with normal, based on prior validation studies; lower scores indicate worse cognitive functioning.

‖Proxy cognition score (range, 0–11): 6–11 is consistent with dementia and 3–5 is cognitive impairment without dementia, and 0–2 is consistent with normal; higher scores indicate worse cognitive functioning.

Table 3 presents the test performance of each algorithm relative to the reference standard. All algorithms that used 3 years of observation achieved similar percentage agreement; however, the Bynum-Standard had significantly higher specificity, resulting in higher PPV. The Bynum-EM algorithm yielded a prevalence estimate (12.7%) that aligned closely with the HRS national dementia prevalence (12.9%), whereas the CCW overestimated prevalence by 1.8% and the standard approach underestimated by 1.6%. The highest PPV and percentage agreement was achieved using the Bynum-Standard 1-year observation algorithm (PPV = 70.3%) at the cost of lower sensitivity and population prevalence estimate. Across all algorithms, sensitivity was generally low (range, 31.3%–56.8%) and specificity was high (range, 92.3%–98.0%). Youden Indices were low (range, 0.293–0.491), largely reflecting the overall low sensitivity of these algorithms. Supplementary Table 2 provides the unweighted sample sizes for each algorithm flag and observation period according to participants' HRS cognitive status of dementia, CIND and normal.

Figure 1 and Supplementary Figure 1 further illustrate the performance characteristics of the Bynum-Standard 1-year algorithm, which had the highest PPV on average in important subgroups. The algorithm's PPV was higher among Black beneficiaries than it is for their White counterparts, driven by its low sensitivity but high specificity (Supplementary Figure 1); that is, when this algorithm identifies a person who is Black as having dementia it is more likely to be a correct diagnosis as measured by HRS. Findings for Hispanics are similar but with broad confidence intervals. Additionally, the PPV is higher for people aged 85 and older and very high for people represented by a proxy. The low PPV for self-respondents, due to particularly low sensitivity, is highly notable, as is the particularly low sensitivity and broad confidence intervals in the group aged 65–74.

Table 4 presents the characteristics of older adults with ADRD in FFS Medicare when claims data are used to identify dementia status. Compared to the national estimates generated using the HRS clinical assessment to identify dementia (in Table 2), all claims algorithms underestimated cases among those with lower educational attainment and rural residence and overestimate comorbidity. They all underestimated cases among minority individuals although a greater proportion of Black beneficiaries are identified using the 1-year algorithms. The 1-year algorithms also identified a greater proportion of people with characteristics of more severe dementia including proxy representation, residing in nursing homes or other facilities, more functional impairments, and higher proportion of individuals who died within 6 months and 1 year of HRS interview. Overall, the characteristics of individuals with dementia more closely approximate the HRS dementia designation using the 1-year algorithms despite having significantly lower population prevalence estimation than HRS.

In the secondary analysis, claims algorithms identifying cognitive impairment, which included the HRS designation of "dementia" or "CIND," had low sensitivity (range, 17.2%–36.6%). PPV (range, 78.6%–89.3%) and specificity (95.5%–99.1%) were higher than for detecting dementia alone, which may make claims useful for ruling out cognitive impairment (Supplementary Table 3). The MCI diagnosis code yielded very low sensitivity (0.03%) for the HRS CIND status (Supplementary Table 4).

Discussion

Revised Medicare claims algorithms have reasonable accuracy for identifying people with dementia for research purposes but investigators should be cognizant of the tradeoffs among the approaches
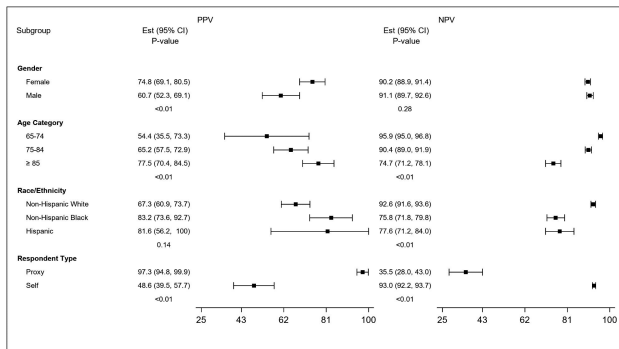
Table 3. Performance of Claims-Based Algorithms Relative to HRS Dementia Status as the Reference Standard

	HRS Dementia Status*			Sensitivity[†] (95% CI)	Specificity[†] (95% CI)	Positive Predictive Value[†] (95% CI)	Negative Predictive Value[†] (95% CI)	Percent Agreement[†]	Youden Index	Algorithm Dementia Prevalence (95%CI)
	Yes	No	Total							
Claims-based dementia status by algorithm										
3-year observation period										
CCW, 3 years n = 5 315				56.8 (53.4–60.1)	92.3 (91.3–93.2)	53.8 (49.4–58.2)	93.1 (92.3–93.9)	87.4%	0.491	14.5% (13.4–15.5)
Yes	503	373	876							
No	363	4 076	4 439							
Total	866	4 449								
Bynum-EM, 3 years n = 5 315				52.2 (48.9–55.5)	93.5 (92.7–94.4)	56.2 (51.5–60.8)	92.5 (91.7–93.3)	87.9%	0.457	12.7% (11.7–13.7)
Yes	466	315	781							
No	400	4 134	4 534							
Total	866	4 449								
Bynum-Standard, 3 years n = 5 315				48.8 (45.5–52.0)	94.9 (94.1–95.7)	60.2 (55.2–65.1)	92.1 (91.3–92.9)	88.6%	0.437	11.1% (10.1–12.1)
Yes	440	262	702							
No	426	4 187	4 613							
Total	866	4 449								
1-year observation period										
Bynum-EM, 1 year n = 5 784				35.4 (31.6–39.3)	97.1 (96.6–97.5)	64.5 (59.1–69.8)	91.0 (90.0–92.0)	89.1%	0.325	7.1% (6.5–7.8)
Yes	351	167	518							
No	576	4 690	5 266							
Total	927	4 857								
Bynum-Standard, 1 year n = 5 784				31.3 (27.5–35.0)	98.0 (97.7–98.4)	70.3 (65.0–75.6)	90.6 (89.5–91.6)	89.4%	0.293	5.8% (5.2–6.3)
Yes	310	117	427							
No	617	4 740	5 357							
Total	927	4 857								

Note: HRS = Health and Retirement Study; CI = confidence interval; CCW = Chronic Condition Data Warehouse.

*Unweighted sample sizes for cross-tabulation between HRS dementia status and the claims-based dementia flag for each algorithm.

[†]Algorithm performance measures and confidence intervals are weighted using the authors' sampling weights, which were constructed to account for characteristics of respondents who did not have linked Medicare data in order to adjust for the complex sampling design. The performance measures reflect national population estimates.

Subgroup	PPV Est (95% CI) P-value	NPV Est (95% CI) P-value
Gender		
Female	74.8 (69.1, 80.5)	90.2 (88.9, 91.4)
Male	60.7 (52.3, 69.1)	91.1 (89.7, 92.6)
	<0.01	0.28
Age Category		
65-74	54.4 (35.5, 73.3)	95.9 (95.0, 96.8)
75-84	65.2 (57.5, 72.9)	90.4 (89.0, 91.9)
≥ 85	77.5 (70.4, 84.5)	74.7 (71.2, 78.1)
	<0.01	<0.01
Race/Ethnicity		
Non-Hispanic White	67.3 (60.9, 73.7)	92.6 (91.6, 93.6)
Non-Hispanic Black	83.2 (73.6, 92.7)	75.8 (71.8, 79.8)
Hispanic	81.6 (56.2, 100)	77.6 (71.2, 84.0)
	0.14	<0.01
Respondent Type		
Proxy	97.3 (94.8, 99.9)	35.5 (28.0, 43.0)
Self	48.6 (39.5, 57.7)	93.0 (92.2, 93.7)
	<0.01	<0.01

Figure 1. Predictive value positive and predictive value negative of the Bynum-standard 1-year algorithm by characteristics. PPV = positive predictive value; NPV = negative predictive value.

they consider. Using the HRS clinical assessment as the reference standard among fee-for-service Medicare beneficiaries, we find that applying an algorithm that mirrors those used for other chronic conditions (ie, 1 year of observation and requiring 1 Part A claim or 2 Part B claims with a revised list of included diagnoses) yields a high PPV, meaning more accurate diagnosis than the commonly used CCW ADRD flag, especially among minority individuals, older adults, and those with more severe disease. However, the tradeoff is lower sensitivity for mild disease and therefore it may be less appropriate for use to generate national population prevalence estimates.

There are several reasons the performance of the revised 1-year algorithm differs from the CCW approach validated in previous research (7,8). The Taylor algorithm validated against the ADAMS HRS subsample, as a gold rather than reference standard, had a higher underlying prevalence of dementia. The algorithm achieved high sensitivity (85%), but lower specificity (89%) and low PPV (56%) (8). The main priority was to maximize sensitivity due to underdiagnosis and biases against using ADRD diagnostic codes related to reimbursement rules, which no longer exist. We find that in the later period, sensitivity prioritization is less needed and we can gain specificity, which limits false positives and improves PPV.

Medicare claims and other administrative data are valuable resources for the rapid identification of people with dementia to target specific populations for population management, quality initiatives, or enrollment in pragmatic clinical trials (6,20,21). In the policy context, they are also used to risk adjust payments (22). When algorithms are used for these purposes, PPV and NPV are more relevant measures of an algorithm's performance than sensitivity and specificity (23,24). We found that the Bynum-Standard 1-year algorithm, in particular, had the highest PPV with excellent specificity and NPV. The greater specificity was also observed compared to the CCW algorithm by requiring 2 claims for services delivered outside of a facility, like other chronic condition algorithms in CCW, which reduces potential misclassification due to reversible symptoms or "rule-out" diagnoses (4,15,20,27). Using the 1-year observation period also decreases the number of people deemed ineligible due to transition to managed care or age below 68.

Using the Bynum-Standard 1-year algorithm, the population of people identified with dementia reflect more severe dementia than with the HRS clinical assessment. Reliance on claims-based algorithms for identifying dementia has an inherent bias whereby people who are sicker, due to primary disease or comorbidity, have more opportunities through contact with the health care system to have their disease identified and billed (20). When compared to HRS-identified dementia cohort, the Bynum-Standard 1-year cohort were older, had more severe functional impairments, and were more likely to use a proxy respondent. For example, the PPV was exceptionally high (97.3%) among those represented by a proxy, which may be due to the high disease prevalence in this subgroup and that they likely receive assistance from care partners in clinic encounters, which may trigger a clinician to bill for dementia. The use of 3 years of claims mitigated this effect substantially.

The higher identification of less severe disease when using 3 years of claims data has the important advantage of leading to approximating national prevalence that more closely aligns with HRS estimates, despite being less accurate. All the 3-year algorithms were within 2% of the HRS population prevalence estimate, while the most specific one (Bynum-EM) was within 0.2%. Recent work by Jain et al. found that having multiple ADRD diagnosis codes when dispersed over time improved the likelihood of having cognitive impairment in HRS (26). In fact, multiple codes dispersed over time were more predictive than the same number of codes in a short time frame suggesting that claims clustered together may suggest a transient confusional state in the absence of ADRD claims over time. They also found that having an ADRD diagnosis combined with a nursing home stay longer than 6 months was associated with having a cognitive impairment (26).

Yet the weakness of all of these algorithms is the under-ascertainment of cases among Black individuals (15,16,25,27,28). Research consistently demonstrates that dementia ascertainment in racial and ethnic minorities is highest based on cognitive assessments and lowest based on claims diagnoses (15,16). The gap between the proportion of Black beneficiaries identified by HRS and claims was smallest using 1-year Bynum-Standard, while all were similarly poor among Hispanic beneficiaries. The reasons for these racial gaps may be related to differences in care-seeking due to stigma (29–31) but may also be due to how we identified dementia within the HRS data. A recent study demonstrates that different methods of using HRS data lead to different estimates of dementia, the best approach by race uses probabilistic modeling (32). We chose to use the approach by Langa, Kabeto, and Weir, which uses directly measured assessments to allow future translation to clinically interpretable pathways. But in all cases, the use of an imperfect reference standard rather than a gold standard may lead to an underestimate of diagnostic accuracy (24). Recent work by Power et al. compared the different dementia ascertainment strategies employed in HRS research using dementia diagnoses in ADAMS as the criterion standard (28). The LKW method had a sensitivity and specificity of 56% and 97%, respectively, with an accuracy of 92%. However, analyses of important subpopulations found that sensitivity was higher among those who were female (68%), Black (84%), had less than a high school education (57%), and had proxy respondents (84%) and specificity was lower in the first 2 groups (84% and 75%, respectively). Conversely, the sensitivity was lower among those who were male (44%), had a high school education or higher (54%), and self-respondents (31%) but specificity was not appreciably different (95%, 98%, and 97%, respectively) (28). Additional work by Gianattasio et al. (32) demonstrated that algorithmic diagnoses such as the LKW method can introduce nonconservative differential bias, particularly when examining racial disparities.

Testing the predictive value of an algorithm in a subpopulation against even an imperfect reference standard allows us to be mindful that the identified population with claims may not be inclusive of all people with dementia. Consistent with other

Table 4. Characteristics of HRS Participants enrolled in FFS Medicare Categorized by Claims-Based Dementia Status

	3 Years			1 Year	
	CCW	Bynum-EM	Bynum-Standard	Bynum-EM	Bynum-Standard
Dementia, sample size	876	781	702	518	427
Dementia, weighted N in millions	3.2M	2.8M	2.4M	1.7M	1.4M
Dementia prevalence, %	14.5%	12.7%	11.1%	7.1%	5.8%
Demographics					
Age in years, mean (SD)	83.7 (7.4)	83.7 (7.3)	83.9 (7.0)	83.7 (7.4)	84.2 (7.1)
Female, %	66.9	66.5	65.1	65.8	68.1
Race/ethnicity*, %					
Non-Hispanic White	81.7	81.9	81.6	80.1	80.2
Non-Hispanic Black	9.9	10.4	10.8	11.5	12.5
Hispanic	6.5	5.9	6.2	6.2	5.7
Education, %					
Less than high school	28.9	28.1	29.7	29.8	32.2
High school and equivalent	32.8	33.1	31.7	33.4	31.5
Some college	19.8	19.7	19.7	17.9	16.9
College graduate and above	18.5	19.2	18.9	18.9	19.4
Residential setting, %					
Community	79.2	77.4	76.7	68.5	66.2
Nursing home or other health care facility	20.8	22.6	23.3	31.5	33.8
Urban/rural, %					
Urban	76.1	77.2	77.2	76.3	76.6
Rural	23.9	22.8	22.8	23.7	23.4
Interview respondent, %					
Self	70.2	68.1	65.6	59.1	55.4
Proxy	29.8	31.9	34.4	40.9	44.6
Functional status					
ADL limitations, %					
0	32.5	32.9	31.1	26.1	23.5
1 or 2	25.2	22.9	24.2	23.9	24.2
3 or more	42.3	44.2	44.7	50.0	52.3
IADL limitations, %					
0	36.6	36.2	33.6	27.5	23.7
1 or 2	23.1	21.7	21.9	20.0	18.6
3 or more	40.3	42.1	44.5	52.5	57.7
Self-reported health status, %					
Excellent, very good, good	49.1	48.4	48.9	44.4	42.2
Fair or poor	50.9	51.6	51.1	55.6	57.8
CCW chronic condition flags, %					
Acute MI/ischemic heart disease	74.8	75.2	74.2	77.9	78.0
Asthma/COPD	44.4	45.6	44.9	46.0	44.2
Atrial fibrillation	30.1	29.1	29.0	29.1	31.7
Chronic kidney disease	43.3	43.2	42.2	46.1	45.2
Diabetes	47.4	48.1	48.6	47.8	49.3
Heart failure	52.0	51.4	51.7	53.7	55.3
Hip or pelvic fracture/osteoporosis	43.9	44.5	43.7	45.7	48.6
Hypertension	92.7	92.9	93.2	93.8	94.9
Stroke/TIA	37.2	37.3	38.8	40.6	43.0

Table 4. Continued

	3 Years			1 Year	
	CCW	Bynum-EM	Bynum-Standard	Bynum-EM	Bynum-Standard
Died, %					
Within 6 months after interview	6.8	7.7	7.7	10.7	11.8
Within 1 year after interview	16.6	18.3	18.7	20.9	24.6
Cognition scales, mean (SD)					
Self-responders					
TICS-m score[†]	9.5 (4.6)	9.4 (4.6)	9.0 (4.5)	8.8 (4.7)	8.2 (4.5)
Proxy responders					
Proxy cognition score[‡]	9.3 (2.1)	9.4 (2.0)	9.5 (1.9)	9.6 (1.8)	9.6 (1.7)

Notes: ADRD = Alzheimer's disease and related dementias; HRS = Health and Retirement Study; SD = standard deviation; ADL = activities of daily living; IADL = instrumental activities of daily living; CCW = Chronic Condition Warehouse; MI = myocardial infarction; COPD = chronic obstructive pulmonary disease; TIA = transient ischemic attack; TICS-m = modified version of the Telephone Interview for Cognitive Status. The reported percentages are weighted using the authors' sampling weights, which were constructed to account for characteristics of respondents who did not have linked Medicare data in order to adjust for the complex sampling design. The weighted *N* (ie, population estimate in millions) and all percentages reflect national population estimates.

*Race/ethnicity does not add to 100; "other category" is suppressed due to small cell sizes.

[†]TICS-m score (range, 0–27); 0–6 is consistent with dementia, 7–11 is consistent with cognitive impairment without dementia, and 12–27 is consistent with normal, based on prior validation studies; lower scores indicate worse cognitive functioning. Sample sizes for TICS-m score among self-respondents for 1-year and 3-year intervals were 5 302 and 4 861, respectively.

[‡]Proxy cognition score (range, 0–11); 6–11 is consistent with dementia, 3–5 is cognitive impairment without dementia, and 0–2 is consistent with normal; higher scores indicate worse cognitive functioning. Sample sizes for the Proxy Cognition Score among proxy respondents for 1-year and 3-year intervals were 482 and 454, respectively.

studies, we found that, in addition to underrepresentation of racial/ethnic minorities, people with low education attainment are underrepresented (15,32,33). There is the potential to introduce health inequities or exacerbate health disparities if caution is not exercised when using claims data to study, payment, or delivery of care for people with dementia (27,34). Reasons for this are complex but include fewer medical encounters and missed or delayed dementia diagnosis resulting in a lower likelihood of racial/ethnic minorities having a dementia diagnosis on a claim (6,8,20,35,36). Therefore, providers and researchers who utilize this claims-based algorithm for dementia care programs should be mindful of ensuring there is a pathway to equitable care for these vulnerable populations.

Another nuanced point about understanding the performance of these claims algorithms is to recognize that this study presents performance for a nationally representative cohort with an underlying prevalence of disease of 12.9%. However, our findings should be interpreted in the context of several additional limitations. First, the performance characteristics of these algorithms may not be generalizable to Medicare Advantage claims, which were not available for this study. Reported dementia prevalence in Medicare Advantage was 5.6% based on the CCW algorithm with a 1-year observation period (37). Next, it is important to consider how coding practices vary across settings of care, specialties, and regions and the influence it may have on identifying dementia diagnosis and interpretation of findings. Predictive values also depend directly on disease prevalence, which will vary from one setting to another affecting the performance of the algorithm (24). Settings with higher dementia prevalence will yield higher predictive values (24,38). Many studies conducted today occur in environments enriched with people living with dementia, such as in nursing homes, home care, or assisted living. Our findings can be used to estimate algorithm performance when applied to these settings of care. In hypothetical settings with dementia prevalence of 20%, 30%, or 50%, PPV of Bynum-Standard algorithm is 76%, 87%, and 94%, respectively. Third, we elected to exclude antidementia medications from our algorithms. Requiring Medicare Part D further limits generalizability. However, Reuben et al. (39) reported that they dropped antidementia medications from their algorithm because they produced a high false-positive rate.

In summary, we examined the performance characteristics of claims-based dementia algorithms using 3- and 1-year observation periods in a contemporary, nationally representative sample of Medicare FFS beneficiaries from the HRS. The Bynum-Standard 1-year algorithm yielded better performance on PPV and specificity with similar NPV to all other algorithms. This algorithm can be useful for rapid identification of Medicare beneficiaries with dementia, but should not be used to estimate dementia prevalence in populations. The Bynum-EM 3-year algorithm is best used in the context of estimating prevalence. All claims-based approaches have limitations when identifying people from minority backgrounds or have lower educational attainment. These revised algorithms have reasonable accuracy for research purposes but investigators should be cognizant of the tradeoffs among the approaches they consider.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Conflict of Interest

None declared.

References

1. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2021;17(3):327–406. doi:10.1002/alz.12328
2. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2020;16:391–460. doi:10.1002/alz.12068
3. Bynum JPW, Langa KM. Prevalence measurement for Alzheimer's disease and dementia: current status and future prospects. *Paper Commissioned for a Workshop on Challenging Questions about Epidemiology, Care, and Caregiving for People with Alzheimer's Disease and Related Dementias and Their Families*. In: National Academies of Sciences E, and Medicine, ed. The National Academies Press; 2020.
4. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of physician-diagnosed Alzheimer's disease and related dementias in population-based administrative data: a validation study using family physicians' electronic medical records. *J Alzheimers Dis*. 2016;54(1):337–349. doi:10.3233/JAD-160105
5. Gilmore-Bykovskyi AL, Jin Y, Gleason C, et al. Recruitment and retention of underrepresented populations in Alzheimer's disease research: a systematic review. *Alzheimers Dement (N Y)*. 2019;5:751–770. doi:10.1016/j.trci.2019.09.018
6. Bynum JPW, Dorr DA, Lima J, et al. Using healthcare data in embedded pragmatic clinical trials among people living with dementia and their caregivers: state of the art. *J Am Geriatr Soc*. 2020;68(suppl 2):S49–S54. doi:10.1111/jgs.16617
7. Taylor DH Jr, Fillenbaum GG, Ezell ME. The accuracy of Medicare claims data in identifying Alzheimer's disease. *J Clin Epidemiol*. 2002;55(9):929–937. doi:10.1016/s0895-4356(02)00452-3
8. Taylor DH Jr, Østbye T, Langa KM, Weir D, Plassman BL. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. *J Alzheimers Dis*. 2009;17(4):807–815. doi:10.3233/JAD-2009-1099
9. CCW condition algorithm: Alzheimer's disease, related disorders, or senile dementia. Accessed October 28, 2021. https://www.ccwdata.org/web/guest/condition-categories
10. St Clair P, Gaudette É, Zhao H, Tysinger B, Seyedin R, Goldman DP. Using self-reports or claims to assess disease prevalence: it's complicated. *Med Care*. 2017;55(8):782–788. doi:10.1097/MLR.0000000000000753
11. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci*. 2011;66(suppl 1):i162–i171. doi:10.1093/geronb/gbr048
12. Langa KM, Larson EB, Crimmins EM, et al. A comparison of the prevalence of dementia in the United States in 2000 and 2012. *JAMA Intern Med*. 2017;177(1):51–58. doi:10.1001/jamainternmed.2016.6807
13. Langa KM, Plassman BL, Wallace RB, et al. The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*. 2005;25(4):181–191. doi:10.1159/000087448
14. Zissimopoulos JM, Tysinger BC, St Clair PA, Crimmins EM. The impact of changes in population health and mortality on future prevalence of Alzheimer's disease and other dementias in the United States. *J Gerontol B Psychol Sci Soc Sci*. 2018;73(suppl 1):S38–S47. doi:10.1093/geronb/gbx147
15. Chen Y, Tysinger B, Crimmins E, Zissimopoulos JM. Analysis of dementia in the US population using Medicare claims: insights from linked survey and administrative claims data. *Alzheimers Dement (N Y)*. 2019;5:197–207. doi:10.1016/j.trci.2019.04.003
16. Zhu Y, Chen Y, Crimmins EM, Zissimopoulos JM. Sex, race, and age differences in prevalence of dementia in Medicare claims and survey data. *J Gerontol B Psychol Sci Soc Sci*. 2021;76(3):596–606. doi:10.1093/geronb/gbaa083
17. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000;53(12):1258–1267. doi:10.1016/s0895-4356(00)00256-0
18. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
19. Weir D, Langa K. How well can Medicare records identify seniors with cognitive impairment needing assistance with financial management? University of Michigan Retirement Research Center (MRRC) Working Paper, WP 2018-391; 2018; Ann Arbor, MI. Accessed October 28, 2021. https://mrdrc.isr.umich.edu/publications/papers/pdf/wp391.pdf
20. Zhu CW, Ornstein KA, Cosentino S, Gu Y, Andrews H, Stern Y. Misidentification of dementia in Medicare claims and related costs. *J Am Geriatr Soc*. 2019;67(2):269–276. doi:10.1111/jgs.15638
21. Lee E, Gatz M, Tseng C, et al. Evaluation of Medicare claims data as a tool to identify dementia. *J Alzheimers Dis*. 2019;67(2):769–778. doi:10.3233/JAD-181005
22. Johnston KJ, Bynum JPW, Joynt Maddox KE. The need to incorporate additional patient information into risk adjustment for Medicare beneficiaries. *JAMA*. 2020;323(10):925–926. doi:10.1001/jama.2019.22370
23. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65(3):343–349.e2. doi:10.1016/j.jclinepi.2011.09.002
24. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537–E544. doi:10.1503/cmaj.121286
25. Thunell J, Ferido P, Zissimopoulos J. Measuring Alzheimer's disease and other dementias in diverse populations using Medicare claims data. *J Alzheimers Dis*. 2019;72(1):29–33. doi:10.3233/JAD-190310
26. Jain S, Rosenbaum PR, Reiter JG, et al. Using Medicare claims in identifying Alzheimer's disease and related dementias. *Alzheimers Dement*. 2020. doi:10.1002/alz.12199. Online ahead of print.
27. Gianattasio KZ, Prather C, Glymour MM, Ciarleglio A, Power MC. Racial disparities and temporal trends in dementia misdiagnosis risk in the United States. *Alzheimers Dement (N Y)*. 2019;5:891–898. doi:10.1016/j.trci.2019.11.008
28. Power MC, Gianattasio KZ, Ciarleglio A. Implications of the use of algorithmic diagnoses or Medicare claims to ascertain dementia. *Neuroepidemiology*. 2020;54(6):462–471. doi:10.1159/000510753
29. Barnes LL, Bennett DA. Alzheimer's disease in African Americans: risk factors and challenges for the future. *Health Aff (Millwood)*. 2014;33(4):580–586. doi:10.1377/hlthaff.2013.1353
30. Holston EC. Stigmatization in Alzheimer's disease research on African American elders. *Issues Ment Health Nurs*. 2005;26(10):1103–1127. doi:10.1080/01612840500280760
31. Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2011;25(3):187–195. doi:10.1097/WAD.0b013e318211c6c9

32. Gianattasio KZ, Wu Q, Glymour MM, Power MC. Comparison of methods for algorithmic classification of dementia status in the health and retirement study. *Epidemiology*. 2019;30(2):291–302. doi:10.1097/EDE.0000000000000945

33. Gianattasio KZ, Ciarleglio A, Power MC. Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the US health and retirement study. *Epidemiology*. 2020;31(1):126–133. doi:10.1097/EDE.0000000000001101

34. Zuckerman IH, Ryder PT, Simoni-Wastila L, et al. Racial and ethnic disparities in the treatment of dementia among Medicare beneficiaries. *J Gerontol B Psychol Sci Soc Sci*. 2008;63(5):S328–S333. doi:10.1093/geronb/63.5.s328

35. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer Dis Assoc Disord*. 2009;23(4):306–314. doi:10.1097/WAD.0b013e3181a6bebc

36. Akushevich I, Yashkin AP, Kravchenko J, Ukraintseva S, Stallard E, Yashin AI. Time trends in the prevalence of neurocognitive disorders and cognitive impairment in the United States: the effects of disease severity and improved ascertainment. *J Alzheimers Dis*. 2018;64(1):137–148. doi:10.3233/JAD-180060

37. Jutkowitz E, Bynum JPW, Mitchell SL, et al. Diagnosed prevalence of Alzheimer's disease and related dementias in Medicare Advantage plans. *Alzheimers Dement (Amst)*. 2020;12(1):e12048. doi:10.1002/dad2.12048

38. Wilkinson T, Ly A, Schnier C, et al; 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

39. Reuben DB, Hackbarth AS, Wenger NS, Tan ZS, Jennings LA. An automated approach to identifying patients with dementia using electronic medical records. *J Am Geriatr Soc*. 2017;65(3):658–659. doi:10.1111/jgs.14744