Title

External Validation of the eRADAR Risk Score for Detecting Undiagnosed Dementia in Two Real-World Healthcare Systems

Permalink

https://escholarship.org/uc/item/6bg9384v

Journal

Journal of General Internal Medicine, 38(2)

ISSN

0884-8734

Authors

Coley, R Yates
Smith, Julia J
Karliner, Leah
et al.

Publication Date

2023-02-01

DOI

10.1007/s11606-022-07736-6

Peer reviewed

External Validation of the eRADAR Risk Score for Detecting Undiagnosed Dementia in Two Real-World Healthcare Systems

R. Yates Coley, PhD[1,2] , Julia J. Smith, MS[1], Leah Karliner, MD MAS[3,4],
Abisola E. Idu, MS MPH[1], Sei J. Lee, MD MAS[5,6], Sharon Fuller, BA[1], Rosemary Lam, BA[4],
Deborah E. Barnes, PhD MPH[5,7,8,9], and Sascha Dublin, MD PhD[1,10]

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; [2]Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA; [3]Multiethnic Health Equity Research Center, University of California San Francisco, San Francisco, USA; [4]Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA; [5]San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA; [6]Division of Geriatrics, Department of Medicine, University of California San Francisco, San Francisco, CA, USA; [7]Weill Institute for Neurosciences, Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, USA; [8]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA; [9]Osher Center for Integrative Health, University of California, San Francisco, San Francisco, CA, USA; [10]Department of Epidemiology, University of Washington, Seattle, WA, USA.

BACKGROUND: Fifty percent of people living with dementia are undiagnosed. The electronic health record (EHR) Risk of Alzheimer's and Dementia Assessment Rule (eRADAR) was developed to identify older adults at risk of having undiagnosed dementia using routinely collected clinical data.

OBJECTIVE: To externally validate eRADAR in two real-world healthcare systems, including examining performance over time and by race/ethnicity.

DESIGN: Retrospective cohort study

PARTICIPANTS: 129,315 members of Kaiser Permanente Washington (KPWA), an integrated health system providing insurance coverage and medical care, and 13,444 primary care patients at University of California San Francisco Health (UCSF), an academic medical system, aged 65 years or older without prior EHR documentation of dementia diagnosis or medication.

MAIN MEASURES: Performance of eRADAR scores, calculated annually from EHR data (including vital signs, diagnoses, medications, and utilization in the prior 2 years), for predicting EHR documentation of incident dementia diagnosis within 12 months.

KEY RESULTS: A total of 7631 dementia diagnoses were observed at KPWA (11.1 per 1000 person-years) and 216 at UCSF (4.6 per 1000 person-years). The area under the curve was 0.84 (95% confidence interval: 0.84–0.85) at KPWA and 0.79 (0.76–0.82) at UCSF. Using the 90th percentile as the cut point for identifying high-risk patients, sensitivity was 54% (53–56%) at KPWA and 44% (38–51%) at UCSF. Performance was similar over time, including across the transition from International Classification of Diseases, version 9 (ICD-9) to ICD-10 codes, and across racial/ethnic groups (though small samples limited precision in some groups).

CONCLUSIONS: eRADAR showed strong external validity for detecting undiagnosed dementia in two health systems with different patient populations and differential availability of external healthcare data for risk calculations. In this study, eRADAR demonstrated generalizability from a research sample to real-world clinical populations, transportability across health systems, robustness to temporal changes in healthcare, and similar performance across larger racial/ethnic groups.

KEY WORDS: decision support tools; dementia; early diagnosis; electronic health record data; prediction.

INTRODUCTION

In the USA, approximately 6.2 million older adults are currently living with dementia.[1,2] Prior studies estimate that only about half have received a diagnosis.[3,4] Individuals who lack a diagnosis may not get the support they need, leaving them vulnerable to potential harms including fragmented medical care, preventable illnesses, safety problems such as car and firearm accidents, and financial abuse.[5–11] Growing awareness of this problem has motivated appeals for large-scale, low-cost strategies to improve dementia detection.[4,12,13] To date, the United States Preventive Services Task Force has concluded that there is insufficient evidence on the benefits and harms to recommend universal dementia screening for adults age 65 and older.[14,15] A targeted screening approach focused on those at elevated risk of dementia may be more effective and feasible for addressing the problem of undiagnosed dementia.

We previously developed an electronic health record (EHR)–based tool that uses routinely collected clinical data to identify patients with increased risk of having undiagnosed dementia, who could potentially be targeted for assessment.[16] The EHR Risk of Alzheimer's and Dementia Assessment Rule (eRADAR) was developed using data collected from 1994 to 2015 for the Adult Changes in Thought (ACT) study, a prospective cohort study of dementia embedded within Kaiser Permanente Washington (KPWA).[17] eRADAR's predictive performance has not been evaluated outside the original sample. Moreover, the ACT cohort was over 90% white, and diagnosis data used to develop eRADAR came from International Classification of Diseases, version 9 (ICD-9) codes; the USA has since transitioned to version 10 (ICD-10). There is a need for robust external validation to assess the accuracy of eRADAR in diverse populations and contemporary practice.

This study externally validated the eRADAR screening tool in two real-world health systems:[18–20] KPWA, an integrated health system, and primary care practices at the University of California San Francisco Health system (UCSF), an academic medical center. This validation study addressed four aspects of eRADAR's performance: (1) generalizability from a research population (ACT) to a real-world clinical population (KPWA); (2) transportability to a non-integrated healthcare system (UCSF) which, like most US healthcare systems, receives little information about care occurring outside of that system; (3) performance over time, including across the 2015 transition from ICD-9 coding to ICD-10 and other temporal changes in clinical practice; and (4) performance across racial/ethnic groups, to ensure that implementation does not exacerbate existing inequities in dementia diagnosis and treatment.[1,21–25]

METHODS

Setting

External validation of eRADAR was performed in two health systems—KPWA and UCSF—and compared to results from internal validation in the original study population (ACT). KPWA is a not-for-profit integrated health system providing insurance coverage and medical care (including primary and specialty care) to about 700,000 members in Washington, including 98,000 Medicare Advantage beneficiaries. UCSF Health is a not-for-profit academic medical system providing primary and specialty care. Three UCSF primary care practices were included in this study: Division of General Internal Medicine, Women's Health, and Lakeshore Family Medicine. These practices deliver medical care to about 48,000 patients, 25% with Medicare or Medicare Advantage insurance.

KPWA and UCSF both utilize the Epic EHR system to record clinical information including encounters, diagnoses, procedures, and medication orders. Epic was deployed at KPWA in 2005 and at UCSF in 2011. KPWA also maintains a research virtual data warehouse with nearly complete capture of outside encounters, diagnoses, and medication dispensings,

which is possible because of its role as an insurance provider; such data are not available in most other health systems,[26] including UCSF, making it especially important to evaluate eRADAR in a more typical setting with data limited to encounters within that health system.

Institutional Review Boards at UCSF and KPWA approved the study procedures and granted waivers of consent and HIPAA authorization.

Study Design

This retrospective validation study was designed to reflect how eRADAR would be implemented in typical healthcare settings. We envision that eRADAR scores would be calculated periodically for patients without a dementia diagnosis, and patients identified as "high risk" would be recommended for cognitive assessment. For this validation analysis, we calculated eRADAR scores annually on January 1 of each study year (2010–2020 at KPWA and 2014–2019 at UCSF) in older patients without dementia based on clinical data in the prior 2 years. We then followed patients to identify incident dementia diagnoses in the subsequent 12 months. People could be included in the sample for multiple years. Predictive performance was evaluated at the person-year level (rather than the person level) to correspond to this design.

Study Population

Validation analyses included all patients aged 65 years or older who met site-specific criteria for adequate utilization with the health system (to ensure availability of clinical data), did not have a dementia diagnosis or dementia medication in the prior 2 years, and were not on hospice care. In order to study eRADAR's validity in a population without documented memory problems, we excluded from the analysis patients with diagnoses in the prior 2 years of amnestic disorder/ memory loss, mild cognitive impairment (MCI), and post-stroke cognitive impairment. Diagnosis codes, medications, and site-specific utilization criteria for inclusion and exclusion are provided in Supplementary Table S1.

KPWA members participating in the ACT study, whose data may have been used in eRADAR model development,[16] were not excluded because patient identifiers were not available under that project's Institutional Review Board approval. Impact on validation was likely negligible; the original eRADAR study only included visits from 1461 ACT study participants between 2010 and 2015, which comprise 1.1% of members in the KPWA validation sample for this study.

eRADAR Prediction Model

eRADAR was originally developed in a randomly selected *training set* (70% of the ACT sample) and internally validated in a testing set (remaining 30%). eRADAR includes 31 predictors of undiagnosed dementia such as demographic characteristics and diagnoses, medications, vital signs, and

healthcare utilization from the prior 2 years (Table S2 and Table 1). For this external validation study, EHR data were extracted for the 2-year period before January 1 of the index year. Diagnoses of comorbid medical conditions were defined using ICD-9 codes recommended by Elixhauser[27] or Charlson[28] and ICD-10 conversions recommended by Quan et al.[29] Details on diagnosis codes and predictor definitions are provided in Table S2.

Outcome: Undiagnosed Dementia

eRADAR was developed to predict risk of undiagnosed dementia, which was identified for ACT participants via formal assessment at biennial study visits.[17] In real-world clinical practice, cognitive screening assessments are not routinely done, so there is no comparable measure of undiagnosed dementia available. These validation analyses used incident dementia diagnosis within 12 months of January 1 of the index year as a proxy for undiagnosed dementia. (See Table S3 for ICD-9/10 diagnosis codes.) Given that dementia diagnoses usually occur relatively late in the disease process,[30] we hypothesized undiagnosed dementia was likely present at the start of the year in which it was diagnosed. For sensitivity analysis, we considered incident dementia diagnosis within 18 months to observe more events and more precisely estimate performance. Additional sensitivity analyses validated eRADAR for a composite outcome of incident dementia or MCI diagnosis within 12 months, since some providers may initially assign an MCI diagnosis when dementia is, in fact, present.[31]

Outcome observation was censored for death or, at KPWA, health plan disenrollment. At KPWA, deaths were identified through patient health records, insurance enrollment records, and state mortality records. At UCSF, deaths were identified

Table 1 Summary of Patient Characteristics by Sample

Characteristic	ACT sample (initial model development)[*]	KPWA validation sample	UCSF validation sample
	N = 16,138 visits[†]	N = 688,599 person-years[†]	N = 47,348 person-years[†]
	# (%)	# (%)	# (%)
Age, mean (SD)	79.9 (6.6)	73.5 (7.3)	73.7 (7.2)
Female	9,721 (60.2)	382,523 (55.6)	27,686 (58.5)
Race/ethnicity			
American Indian or Alaskan Native	27 (0.2)	9291 (1.3)	121 (0.3)
Asian or Asian American	597 (3.7)	43,819 (6.4)	14,677 (31.0)
Black or African American	582 (3.6)	19,671 (2.9)	3623 (7.7)
Hispanic or Latinx	126 (0.8)	22,106 (3.2)	3517 (7.4)
Multiple races[‡]	N/A	10,064 (1.5)	N/A
Native Hawaiian or other Pacific Islander	1 (< 0.1)	3015 (0.4)	417 (0.9)
Other race[§]	299 (1.9)	11,062 (1.6)	2135 (4.5)
White non-Hispanic	14,495 (89.8)	568,720 (82.6)	22,477 (47.5)
No race or ethnicity recorded	11 (<0.1)	20,939 (3.0)	491 (1.0)
Preferred language			
Chinese-Cantonese	N/A[‖]	1973 (0.3)	2766 (5.8)
Chinese-Mandarin	N/A	1232 (0.2)	1489 (3.1)
English	N/A	652,236 (94.7)	37,953 (80.2)
Spanish	N/A	766 (0.1)	1315 (2.8)
All other languages	N/A	21,758 (3.2)	3808 (8.0)
No preference specified	N/A	10,634 (1.5)	17 (< 0.1)
Insurance type			
Medicare Advantage	N/A[‖]	672,931 (97.7)[¶]	7537 (15.9)[#]
Medicare fee-for-service	N/A	N/A[**]	30,108 (63.6)
Medicaid	N/A	14 (< 0.1)	1469 (3.1)
Commercial	N/A	275,426 (40.0)	8023 (16.9)
No insurance	N/A	N/A[**]	211 (0.4)
Elixhauser comorbidity score,[27] mean (sd)	N/A[††]	3.4 (6.6)	3.8 (6.3)

ACT Adult Changes in Thought, IQR interquartile range, KPWA Kaiser Permanente Washington, SD standard deviation, UCSF University of California San Francisco

[*]*ACT sample includes 11,431 visits in the training set (used for initial model estimation) and 4707 visits in the testing set (used for internal validation). Because training/testing assignments were random, characteristics of the training and testing sets were similar and reflect the overall sample*

[†]*Individual people may contribute multiple observations to validation sets; i.e., an ACT study participant may have multiple study visits and KPWA and UCSF members may contribute multiple person-years*

[‡]*Multiple races, or multiracial, is not an option for patient self-report at UCSF or in the ACT study. At KPWA, patients can report up to 5 races. Patients with "multiple races" or more than one race indicated were included in this category. Because patients at KPWA can specify multiple races, the sum of percentages across all racial/ethnic categories will be greater than 100%*

[§]*"Other" race is an option for patient self-report*

[‖]*Preferred language and type of insurance coverage score information was not available for the ACT study*

[¶]*Some KPWA patients may have more than one insurance plan listed in their medical records, such that the sum of patient-years across insurance types will be greater than the number of person-years in the sample*

[#]*For UCSF patient-years, the primary insurance payor is indicated*

[**]*The KPWA validation sample was restricted to patients enrolled in a KPWA insurance plan and, as a result, does not include patients with Medicare Fee-for-service insurance or without insurance*

[††]*Elixhauser comorbidity score information was not available for the ACT study*

through patient health records, which only include deaths at UCSF hospitals or for which the healthcare team is notified.

Measuring eRADAR Performance

We examined measures of performance that reflected how health systems would use eRADAR to identify high-risk patients to target for dementia assessment. To select an eRADAR cut point above which patients are classified as "high risk," health system leaders consider whether that cut point accurately identifies people with undiagnosed dementia (sensitivity) while limiting unneeded evaluations for people without undiagnosed dementia (specificity). At each threshold, the intensity and cost of an intervention should be appropriate for the rate of undiagnosed dementia among those flagged as "high risk" (positive predictive value [PPV]). We considered the 99th, 95th, 90th, 85th, and 75th percentiles of the eRADAR score because these would be realistic, feasible cut points that a healthcare system might use. Performance was also evaluated using area under the curve (AUC), which summarizes sensitivity and specificity across all possible thresholds.[32]

eRADAR was also validated within subgroups defined by race/ethnicity. Race/ethnicity information is collected by both health systems via patient self-report at clinical visits. Neither race nor ethnicity is a predictor in eRADAR. The performance of eRADAR within racial/ethnic subgroups was evaluated for the 18-month incident dementia diagnosis to increase statistical power to detect differences across subgroups. "Fairness" was assessed on the basis of similar AUC, sensitivity (also known as *equalized opportunity*), and PPV (*predictive parity*) across race/ethnicity.[33]

Performance estimates were adjusted for censoring due to health plan disenrollment or death using inverse probability weighting.[34] Analytic details are provided in the online supplement.

RESULTS

Analyses included 688,599 person-years among 129,315 patients at KPWA and 47,348 person-years among 13,444 patients at UCSF. Compared to the ACT cohort in which eRADAR was originally developed, external validation samples were younger and more racially/ethnically diverse and had fewer comorbidities and less healthcare utilization (Tables 1 and 2). The UCSF sample had a greater proportion of observations from Asian/Asian American, Black/African American, and Hispanic/Latinx patients than the KPWA sample. Medicare fee-for-service insurance coverage was also more common at UCSF than KPWA.

eRADAR scores, that is, estimates of the risk of undiagnosed dementia, were generally low for both validation samples (median [interquartile range] = 1.0% [0.6–1.9%] for both) and about half of those observed in the ACT cohort (2.1% [1.2–3.8%]). Characteristics, predictors, and eRADAR scores

of external validation samples were similar across study years (Table S4).

Incident dementia diagnoses were recorded for 7631 KPWA patients (a rate of 11.1 events per 1000 person-years) and 216 UCSF patients (4.6 events per 1000 person-years). As expected, given differences in sample characteristics and study design, a higher rate of undiagnosed dementia was observed in the ACT sample: 31 per 1000 biennial visits or 15 per 1000 person-years. Incident dementia rates in the UCSF sample declined from 6.0 per 1000 person-years in 2014 to 3.5 per 1000 person years in 2019 (Table S5). Rates at KPWA ranged between 9.3 and 14.4 events per 1000 person-years with no distinctive time trend.

Figure 1 shows receiver operating characteristic curves from the original internal validation sample (ACT) and current external validation samples (KPWA and UCSF). AUC was greater for KPWA (0.84 [95% confidence interval {CI}: 0.84–0.85]) than UCSF (0.79 [0.76–0.82]). AUC 95% CIs for both external validation samples overlapped with that of the ACT internal validation sample (0.81 [0.78–0.84]).

eRADAR sensitivity was highest in the KPWA sample and similar between UCSF and ACT (Table 2). For example, classifying those with eRADAR scores above the 90th percentile as high-risk captures 54% (95% CI: 53–56%) of KPWA person-years with incident dementia diagnosis, 44% (38–51%) of UCSF person-years with dementia diagnosis, and 36% (28–44%) of ACT visits with undiagnosed dementia.

eRADAR PPV was greater in the ACT internal validation sample than the external validation samples (Table 3), as was expected because PPV is strongly influenced by the outcome rate.[35] To quantify how eRADAR could improve identification of high-risk individuals for dementia evaluation, we compared outcome rates in the entire sample (which represents the PPV of universal screening) to rates among those designated as high risk at a given cut point (PPV of eRADAR). In the ACT testing set, visits with eRADAR scores above the 90th percentile were 3.7 times more likely to have undiagnosed dementia than the average visit. Similarly, KPWA and UCSF person-years with eRADAR scores above the 90th percentile were, respectively, 4.3 and 4.4 times more likely to receive an incident dementia diagnosis within the next year than the average person-year.

eRADAR performance in external validation samples was similar across years (Figure S1) and across the ICD-9/10 transition (Fig. 2). AUC at KPWA was 0.84 during ICD-9 years and 0.85 during ICD-10; AUC at UCSF was 0.79 and 0.78 in ICD-9 and ICD-10 years, respectively (Figure S2). PPV showed small, not clinically meaningful, variability year to year, likely because PPV is sensitive to small changes in outcome rate when outcomes are relatively rare, as seen in these samples.

In sensitivity analyses, changes in outcome definition and ascertainment period did not meaningfully affect performance (Tables S6-S7). Changes in PPV followed the expected

Table 2 Prevalence of eRADAR Predictors by Sample

Characteristic	ACT sample (initial model development)[*]	KPWA validation sample	UCSF validation sample
	N = 16,138 visits[**]	N = 688,599 person-years[**]	N = 47,348 person-years[**]
	# (%)	# (%)	# (%)
Diagnoses, past 2 years			
Congestive heart failure	1978 (12.3)	52,083 (7.6)	2741 (5.8)
Cerebrovascular disease	1833 (11.4)	50,341 (7.3)	2818 (6.0)
Diabetes, any	2342 (14.5)	135,194 (19.6)	12,176 (25.7)
Diabetes, complex	1153 (7.1)	85,568 (12.4)	5098 (10.8)
Chronic pulmonary disease	3013 (18.7)	128,206 (18.6)	6952 (14.7)
Hypothyroidism	2121 (13.1)	90,216 (13.1)	7616 (16.1)
Renal failure	1654 (10.2)	116,400 (16.9)	6011 (12.7)
Lymphoma	174 (1.1)	6311 (0.9)	774 (1.6)
Solid tumor w/o metastases	3755 (23.3)	118,194 (17.2)	11,411 (24.1)
Rheumatoid arthritis	903 (5.6)	34,496 (5.0)	1716 (3.6)
Weight loss	94 (0.6)	14,674 (2.1)	2825 (6.0)
Fluid and electrolyte disorders	2013 (12.5)	75,294 (10.9)	5175 (10.9)
Blood loss anemia	691 (4.3)	7384 (1.1)	486 (1.0)
Bipolar disorder and psychoses	343 (2.1)	9119 (1.3)	976 (2.1)
Depression	2364 (14.6)	120,961 (17.6)	6530 (13.8)
Traumatic brain injury	233 (1.4)	22,137 (3.2)	1021 (2.2)
Tobacco use (past or current)	1156 (7.2)	108,351 (15.7)	2137 (4.5)
Atrial fibrillation	2377 (14.7)	77,279 (11.2)	4449 (9.4)
Gait abnormality	1379 (8.5)	45,656 (6.6)	2764 (5.8)
Vital signs			
BMI < 18.5 kg/m^2	265 (1.6)	7263 (1.1)	1184 (2.5)
BMI ≥30 kg/m^2	3576 (22.2)	211,240 (30.7)	9627 (20.3)
High blood pressure (systolic ≥ 140 mmHg or diastolic ≥ 90 mmHg)	6475 (40.1)	145,509 (21.1)	14,338 (30.3)
Healthcare utilization, past 2 years			
≥ 1 outpatient primary care visit	15,939 (98.8)	667,420 (96.9)	47,133 (99.5)
≥ 1 emergency department visit	4325 (26.8)	131,984 (19.2)	10,222 (21.6)
Home health services	1469 (9.1)	38,608 (5.6)	2011 (4.2)
≥ 1 physical therapy visit	6083 (37.7)	197,226 (28.6)	18,117 (38.3)
≥ 1 cognitive testing referral [‡]	367 (2.3)	5447 (0.8)	54 (0.1)
Medications, past 2 years[‡]			
Antidepressants	2237 (13.9)	140,451 (20.4)	10,019 (21.2)
Sleep aids	4294 (26.6)	142,361 (20.7)	14,983 (31.6)
eRADAR risk score, median (IQR)	2.1% (1.2–3.8%)	1.0% (0.6–1.9%)	1.0% (0.6–1.9%)

ACT Adult Changes in Thought, BMI body mass index, KPWA Kaiser Permanente Washington, SD standard deviation, UCSF University of California San Francisco

[*]*ACT sample includes 11,431 visits in the training set (used for initial model estimation) and 4707 visits in the testing set (used for internal validation). Because training/testing assignments were random, characteristics of the training and testing sets were similar and reflect the overall sample*

[**]*Individual people may contribute multiple observations to validation sets, i.e., an ACT study participant may have multiple study visits and KPWA and UCSF members may contribute multiple person-years*

[‡]*In ACT and KPWA samples, defined as a visit in the Speech, Language and Learning Department, which provides the vast majority of in-depth cognitive evaluations for KPWA patients; at UCSF, defined as a referral for neuropsychological testing*

[§]*In ACT and KPWA samples, defined by medication dispensing; at UCSF, defined as medication order (because dispensing data are not available in the EHR)*

pattern (higher PPV with higher outcome incidence) and did not indicate unanticipated variation in performance.

eRADAR also performed similarly across racial/ethnic groups (Fig. 3, Table S8), though there was wider variability in estimates for subgroups at UCSF due to smaller sample sizes.

DISCUSSION

eRADAR uses routinely collected clinical data to predict risk of undiagnosed dementia in older patients, with the goal of identifying high-risk patients who could potentially be targeted for dementia assessment. In this study, eRADAR demonstrated strong external validity in two diverse health

systems, including evaluation of temporal trends and racial/ethnic differences.

Our findings affirm the generalizability of eRADAR from a research sample to real-world clinical populations. Volunteers for research studies are typically not representative of general patient populations,[36–38] and, as shown in our data, ACT participant characteristics do not reflect those of KPWA members overall. We found that eRADAR accurately predicted dementia risk in a wider sample of KPWA members.

This study also demonstrated eRADAR's transportability to a new setting that better represents health systems in which the prediction model is likely to be implemented.[19,20] eRADAR was developed using data from KPWA, an integrated healthcare system that has nearly complete capture of external care, unlike most US healthcare systems. UCSF is more
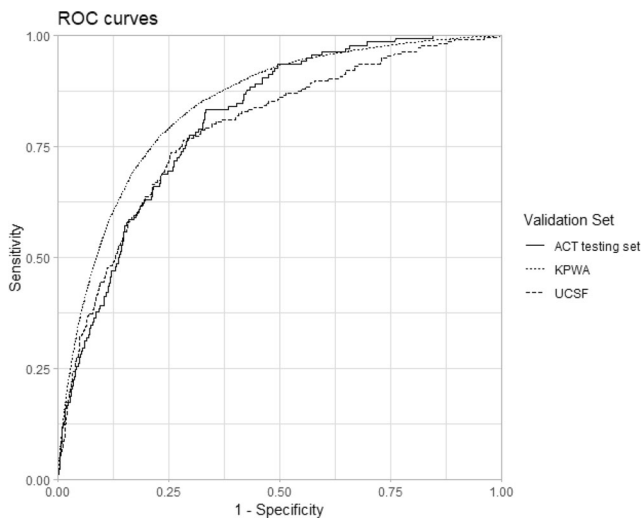
Figure 1 Receiver operating characteristic curves for eRADAR prediction of 12-month incident dementia diagnosis in ACT testing set (solid line), KPWA validation sample (dotted line), and UCSF validation sample (dashed line).

typical in that it has easy access to EHR data only for care provided within the health system, and many patients receive additional care in other settings. UCSF also serves a patient population with more socioeconomic and racial/ethnic diversity than KPWA. Successful external validation of eRADAR at UCSF suggests that eRADAR may accurately predict undiagnosed dementia risk in a variety of healthcare settings and populations.

Our findings highlight the need to match clinical prediction models with appropriate interventions based on their performance characteristics. eRADAR shows high discrimination, sensitivity, and specificity in both healthcare settings examined, but the PPV remains low due to a low incidence of dementia diagnosis in the populations served.[39] Because most patients classified as high risk by eRADAR will not have dementia, follow-up should not be overly invasive, expensive, or burdensome.[12] Care should be taken to address the potential
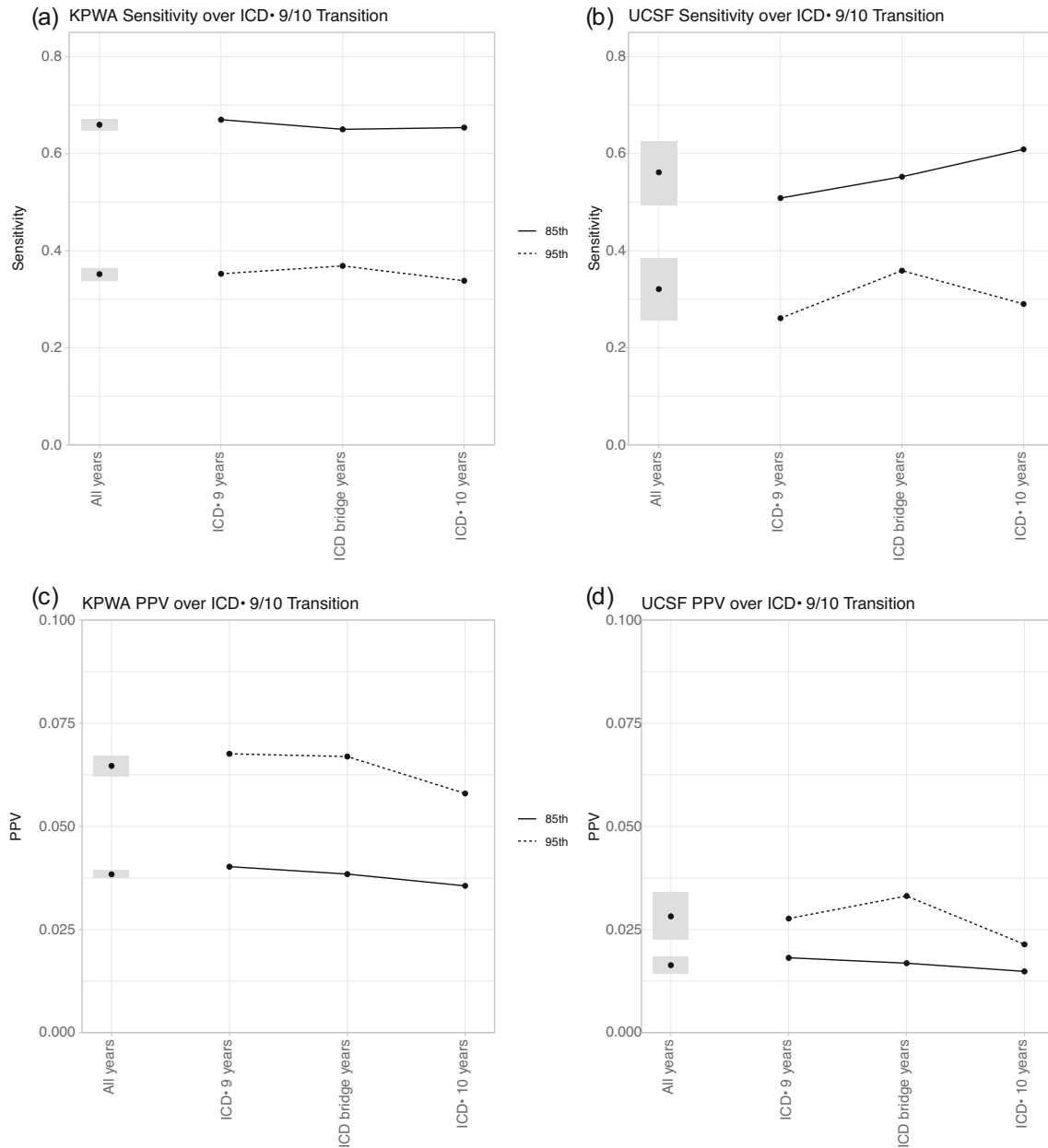
for stigma, anxiety, and increased suicide risk that may result from patients being identified as high risk.[6,7,40,41]

One study limitation is that outcomes were derived from dementia diagnoses indicated in the EHR (perhaps without formal cognitive assessment) and, as such, were susceptible to misclassification, including both under- and overdiagnosis. In particular, validation analyses likely underestimated PPV because dementia is underdiagnosed (by as much as 50%) in routine clinical practice.[3,4] Original estimation of eRADAR benefited from a "gold standard" assessment of dementia status (biennial cognitive screening in a research study), in which it was likely that very few dementia cases were missed.[17] For this large-scale validation study, it was not feasible to conduct universal cognitive screening. As an alternative, we evaluated how well the eRADAR model predicted future dementia diagnoses in the EHR. We are currently planning a randomized pragmatic trial to examine the impact of providing cognitive and functional evaluations to patients with high eRADAR scores. Results from that study will improve estimation of eRADAR's performance among high-risk patients as well as identify factors supporting and barriers to implementation of eRADAR in clinical practice.

Comparing the performance of clinical prediction models across racial and ethnic groups is fundamental to ensuring their use does not exacerbate existing inequities in access to needed healthcare services.[33,42–44] Our analysis indicated that eRADAR provided *equal opportunity* for benefit across race/ethnicity, that is, a similar proportion of patients later diagnosed with dementia were correctly identified as high risk. eRADAR also showed *predictive parity* across racial/ethnic groups, meaning that patients classified as high risk had similar rates of incident dementia diagnoses. A limitation of this analysis is that the precision of performance estimates was lower within subgroups with a small number of incident dementia diagnoses observed, particularly Native Hawaiian and Pacific Islander members at KPWA and all racial/ethnic groups at UCSF except White non-Hispanic and Asian

Table 3 Classification Accuracy (% [95% CI]) of eRADAR Prediction Model by Validation Sample

Measure, risk cut-off percentile	ACT, internal validation	KPWA, external validation	UCSF, external validation
Sensitivity			
$\geq 99^{th}$	6.5 (2.9–10.9)	9.5 (8.7–10.3)	7.1 (3.7–11.0)
$\geq 95^{th}$	22.5 (15.9–29.7)	35.2 (33.9–36.4)	32.1 (25.7–38.5)
$\geq 90^{th}$	36.2 (28.3–44.2)	54.3 (52.9–55.6)	44.4 (37.6–50.9)
$\geq 85^{th}$	47.1 (39.1–55.8)	65.9 (64.7–67.1)	56.2 (49.4–62.6)
$\geq 75^{th}$	65.9 (58.0–73.2)	79.6 (78.6–80.6)	73.6 (67.4–79.4)
Specificity			
$\geq 99^{th}$	99.4 (99.1–99.6)	99.2 (99.2–99.3)	98.9 (98.7–99.0)
$\geq 95^{th}$	96.4 (95.6–97.1)	95.3 (95.2–95.4)	94.9 (94.5–95.2)
$\geq 90^{th}$	91.6 (90.5–92.8)	89.9 (89.8–90.1)	89.8 (89.2–90.3)
$\geq 85^{th}$	87.2 (85.8–88.7)	84.6 (84.4–84.8)	84.4 (83.8–85.0)
$\geq 75^{th}$	78.0 (76.2–79.9)	74.4 (74.2–74.7)	74.4 (73.7–75.2)
PPV			
$\geq 99^{th}$	23.7 (11.4–37.5)	10.1 (9.4–10.9)	2.9 (1.5–4.4)
$\geq 95^{th}$	15.7 (11.2–20.6)	6.5 (6.2–6.7)	2.8 (2.3–3.4)
$\geq 90^{th}$	11.5 (9.1–14.2)	4.8 (4.6–4.9)	2.0 (1.7–2.3)
$\geq 85^{th}$	10.0 (8.3–12.0)	3.8 (3.7–3.9)	1.6 (1.4–1.8)
$\geq 75^{th}$	8.3 (7.3–9.5)	2.8 (2.8–2.9)	1.3 (1.2–1.4)

ACT Adult Changes in Thought, CI confidence interval, KPWA Kaiser Permanente Washington, UCSF University of California, San Francisco
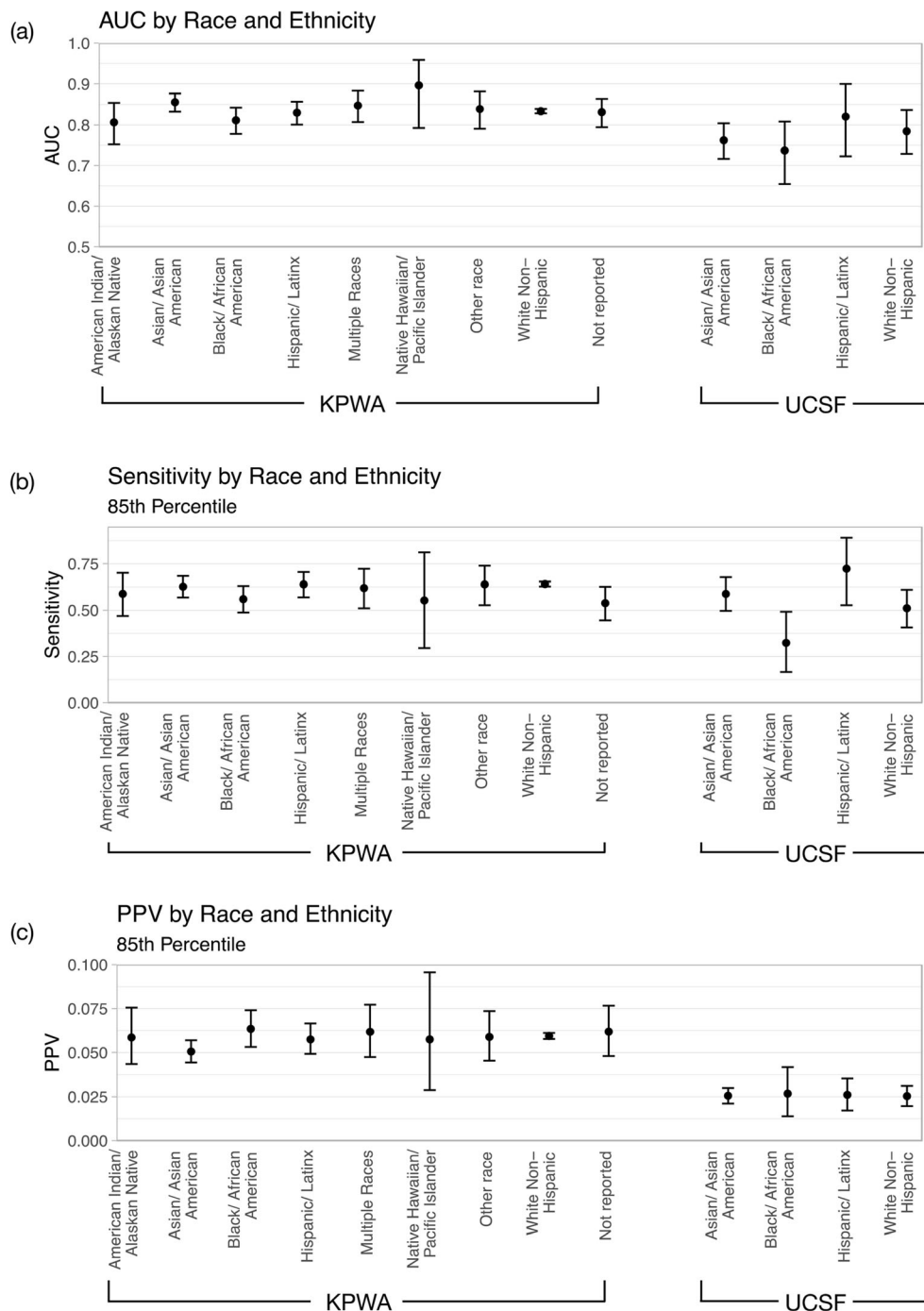
Figure 2 Performance of eRADAR for predicting 12-month incident dementia diagnosis in KPWA and UCSF external validation sets across the International Classification of Disease (ICD) version 9 to 10 transition, measured by sensitivity (a and b) and positive predictive value (PPV, c and d) of eRADAR scores above the 85th (solid line) and 95th (dotted line) percentiles. Shaded regions indicate point-wise 95% confidence intervals. ICD-9 years include all observations with an index year of 2014 or earlier (before the ICD-10 transition). ICD bridge years include observations from 2015 to 2017 for which clinical data in the prior 2 years (used to calculate the eRADAR score) may include ICD-9 diagnosis codes; additionally, incident dementia diagnoses for 2015 observations may include ICD-9 or ICD-10 codes, as the transition occurred on October 1, 2015. ICD-10 years include observations from 2018 and later years where eRADAR predictors and outcomes were defined using ICD-10 diagnosis codes exclusively.

patients (including American Indian and Alaskan Native, Black, Hispanic, and Native Hawaiian and Pacific Islander patients). Additionally, underdiagnosis of dementia is disproportionately common in Black/African American and Hispanic/Latinx patients;[21–23] the potential for differential outcome misclassification is an additional limitation of our validation analyses within racial/ethnic groups. To address these gaps, our trial will prospectively monitor eRADAR accuracy in racial/ethnic groups.

eRADAR performance was also robust to temporal variability, including the ICD-9/10 transition. Temporal validation is essential for clinical prediction models intended for real-world deployment to establish prospective accuracy.[19,20] Performance of models that depend on healthcare utilization (including diagnostic assessments, measuring vital signs, or performing labs and imaging) may be impacted by changes in usual care practices. This study's evaluation of temporal changes in eRADAR's performance was unable to capture the potential impact of the

Figure 3 Performance of eRADAR for predicting 18-month incident dementia diagnosis in racial/ethnic subgroups at KPWA and UCSF, measured by area under the curve (AUC, a) and by sensitivity (b) and positive predictive value (PPV, c) of eRADAR scores above the 85th percentile. Point estimates and 95% confidence intervals (lines) are shown. AT UCSF, there were too few observations and events within person-years with American Indian/Alaskan Native, Native Hawaiian/ Pacific Islander, or other race indicated or without race/ethnicity recorded to evaluate performance within these groups. Multiple races, or multiracial, is not an option for patient self-report at UCSF.

COVID-19 pandemic, including interruptions in healthcare and the shift to telemedicine. External validation at KPWA included incident dementia diagnoses observed in 2020, but pandemic effects are not adequately reflected in eRADAR predictors, which are measured over the preceding 2 years, for any

observations in this study. Ongoing monitoring of performance is needed for any clinical prediction model to verify that accuracy is maintained and is particularly important given the major impact of the COVID-19 pandemic on healthcare utilization including deferred care and telemedicine.

CONCLUSION

This study validated eRADAR, which predicts risk of undiagnosed dementia using EHR data, in two real-world health systems whose patient populations differ from the original development sample. External validation demonstrated eRADAR's generalizability to a non-research sample and transportability to a different US health system. eRADAR's performance was robust to temporal changes and consistent across racial/ethnic groups. eRADAR may be a useful tool to help healthcare providers identify high-risk patients to target for dementia assessment.

Corresponding Author: R. Yates Coley, PhD; Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA (e-mail: Rebecca.Y.Coley@kp.org).

Declarations:

REFERENCES

1. 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021;17(3):327-406. https://doi.org/10.1002/alz.12328.
2. **Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA.** Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). Alzheimers Dement. 2021. https://doi.org/10.1002/alz.12362.
3. **Lang L, Clifford A, Wei L, et al.** Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7(2):e011146. https://doi.org/10.1136/bmjopen-2016-011146.
4. 2019 Alzheimer's disease facts and figures. Alzheimers Dement. 2019;15(3):321-87.
5. **Amjad H, Roth DL, Samus QM, Yasar S, Wolff JL.** Potentially unsafe activities and living conditions of older adults with dementia. J Am Geriatr Soc. 2016;64(6):1223-32. https://doi.org/10.1111/jgs.14164.
6. **Borson S, Frank L, Bayley PJ, et al.** Improving dementia care: the role of screening and detection of cognitive impairment. Alzheimers Dement. 2013;9(2):151-9. https://doi.org/10.1016/j.jalz.2012.08.008.
7. **Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G.** Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. J Alzheimers Dis. 2016;49(3):617-31. https://doi.org/10.3233/jad-150692.
8. **Dreier-Wolfgramm A, Michalowsky B, Austrom MG, et al.** Dementia care management in primary care: current collaborative care models and the case for interprofessional education. Z Gerontol Geriatr. 2017;50(Suppl 2):68-77. https://doi.org/10.1007/s00391-017-1220-8.
9. **Lin PJ, Fillit HM, Cohen JT, Neumann PJ.** Potentially avoidable hospitalizations among Medicare beneficiaries with Alzheimer's disease and related disorders. Alzheimers Dement. 2013;9(1):30-8. https://doi.org/10.1016/j.jalz.2012.11.002.
10. **Betz ME, McCourt AD, Vernick JS, Ranney ML, Maust DT, Wintemute GJ.** Firearms and dementia: clinical considerations. American College of Physicians; 2018. p. 47-49.
11. **Rosen T, Makaroun LK, Conwell Y, Betz M.** Violence in older adults: scope, impact, challenges, and strategies for prevention. Health Aff. 2019;38(10):1630-1637.
12. The Gerontological Society of America. The Gerontological Society of America Workgroup on Cognitive Impairment Detection and Earlier Diagnosis Report and Recommendations. 2015.
13. **Cordell CB, Borson S, Boustani M, et al.** Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. Alzheimers Dement. 2013;9(2):141-50. doi:https://doi.org/10.1016/j.jalz.2012.09.011
14. **Owens DK, Davidson KW, Krist AH, et al.** Screening for cognitive impairment in older adults: US Preventive Services Task Force recommendation statement. JAMA. 2020;323(8):757-763. https://doi.org/10.1001/jama.2020.0435.
15. **Patnode CD, Perdue LA, Rossom RC, et al.** Screening for cognitive impairment in older adults: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA. 2020;323(8):764-785. https://doi.org/10.1001/jama.2019.22258.
16. **Barnes DE, Zhou J, Walker RL, et al.** Development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. J Am Geriatr Soc. 2020;68(1):103-111. https://doi.org/10.1111/jgs.16182.
17. **Kukull WA, Higdon R, Bowen JD, et al.** Dementia and Alzheimer disease incidence: a prospective cohort study. Arch Neurol. 2002;59(11):1737-46. https://doi.org/10.1001/archneur.59.11.1737.
18. **Altman DG, Royston P.** What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-473.
19. **Justice AC, Covinsky KE, Berlin JA.** Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-524. https://doi.org/10.7326/0003-4819-130-6-199903160-00016.
20. **König IR, Malley JD, Weimar C, Diener HC, Ziegler A.** Practical experiences on the necessity of external validation. Stat Med. 2007;26(30):5499-511. https://doi.org/10.1002/sim.3069.
21. **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. https://doi.org/10.1007/s11606-018-4377-y.
22. **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. https://doi.org/10.1016/j.trci.2019.11.008.
23. **Lin PJ, Daly AT, Olchanski N, et al.** Dementia diagnosis disparities by race and ethnicity. Med Care. 2021;59(8):679-686. https://doi.org/10.1097/mlr.0000000000001577.
24. **Cooper C, Tandy AR, Balamurali TB,** Livingston G. A systematic review and meta-analysis of ethnic differences in use of dementia treatment, care, and research. Am J Geriatr Psychiatry. 2010;18(3):193-203. https://doi.org/10.1097/JGP.0b013e3181bf9caf.
25. **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):S328-33. https://doi.org/10.1093/geronb/63.5.s328.
26. **Ross TR, Ng D, Brown JS, et al.** The HMO Research Network Virtual Data Warehouse: a public data model to support collaboration. EGEMS (Washington, DC). 2014;2(1):1049. https://doi.org/10.13063/2327-9214.1049.

27. **Elixhauser A, Steiner C, Harris DR, Coffey RM.** Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8-27. https://doi.org/10.1097/00005650-199801000-00004.

28. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83. https://doi.org/10.1016/0021-9681(87)90171-8.

29. **Quan H, Sundararajan V, Halfon P, et al.** Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130-9. https://doi.org/10.1097/01.mlr.0000182534.19832.83.

30. **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):306-14. https://doi.org/10.1097/WAD.0b013e3181a6bebc.

31. What is mild cognitive impairment, how it differs from dementia & how a diagnosis is made. https://www.dementiacarecentral.com/aboutdementia/othertypes/mci/. Accessed 15 Dec 2021.

32. **Hanley JA, McNeil BJ.** A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148(3):839-43. https://doi.org/10.1148/radiology.148.3.6878708.

33. **Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH.** Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-872.

34. **Robins JM RA, Zhao LP.** Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J Am Stat Assoc. 1995;90:106-121.

35. **Tenny S, Hoffman MR**. Prevalence. https://www.ncbi.nlm.nih.gov/books/NBK430867/. Accessed 15 Dec 2021.

36. **Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A** literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16:495. https://doi.org/10.1186/s13063-015-1023-4.

37. **Gleason CE, Norton D, Zuelsdorff M, et al.** Association between enrollment factors and incident cognitive impairment in Blacks and Whites: data from the Alzheimer's Disease Center. Alzheimers Dement. 2019;15(12):1533-1545. https://doi.org/10.1016/j.jalz.2019.07.015.

38. **Barnes LL, Bennett DA.** Alzheimer's disease in African Americans: risk factors and challenges for the future. Health affairs (Project Hope). 2014;33(4):580-6. https://doi.org/10.1377/hlthaff.2013.1353.

39. **Simon GE, Shortreed SM, Coley RY.** Positive predictive values and potential success of suicide prediction models. JAMA Psychiatry. 2019;76(8):868-869. https://doi.org/10.1001/jamapsychiatry.2019.1516.

40. **Günak MM, Barnes DE, Yaffe K, Li Y, Byers AL.** Risk of suicide attempt in patients with recent diagnosis of mild cognitive impairment or dementia. JAMA Psychiatry. 2021;78(6):659-666. https://doi.org/10.1001/jamapsychiatry.2021.0150.

41. **Schmutte T, Olfson M, Maust DT, Xie M, Marcus SC.** Suicide risk in first year after dementia diagnosis in older adults. Alzheimers Dement. 2022;18(2):262-271.

42. **Matheny M, Israni ST, Ahmed M, Whicher D**. Artificial intelligence in health care: the hope, the hype, the promise, the peril. Natl Acad Med. 2020:94-97.

43. **Rajkomar A, Dean J, Kohane I.** Machine learning in medicine. N Engl J Med. 2019;380(14):1347-1358.

44. **Parikh RB, Teeple S, Navathe AS.** Addressing bias in artificial intelligence in health care. JAMA. 2019;322(24):2377-2378.