Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Neuroepidemiology. Author manuscript; available in PMC 2021 October 19.

Published in final edited form as:

Neuroepidemiology. 2020 ; 54(6): 462–471. doi:10.1159/000510753.

Implications of the use of algorithmic diagnoses or Medicare claims to ascertain dementia

Melinda C. Power[a], **Kan Z. Gianattasio**[a], **Adam Ciarleglio**[b]

[a]Department of Epidemiology, Milken Institute School of Public Health, George Washington University, Washington, DC

[b]Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington, DC

Abstract

Introduction: Formal dementia ascertainment with research criteria is resource-intensive, prompting growing use of alternative approaches. Our objective was to illustrate the potential bias and implications for study conclusions introduced through use of alternate dementia ascertainment approaches.

Methods: We compared dementia prevalence and risk factor associations obtained using criterion-standard dementia diagnoses to those obtained using algorithmic or Medicare-based dementia ascertainment in participants of the baseline visit of the Aging, Demographics, and Memory Study (ADAMS), a Health and Retirement Study (HRS) sub-study.

Results: Estimates of dementia prevalence derived using algorithmic or Medicare-based ascertainment differ substantially from those obtained using criterion-standard ascertainment. Use of algorithmic or Medicare-based dementia ascertainment can, but does not always lead to risk-factor associations that substantially differ from those obtained using criterion-standard ascertainment.

Discussion/Conclusions: Absolute estimates of dementia prevalence should rely on samples with formal dementia ascertainment. Use of multiple algorithms is recommended for risk-factor studies when formal dementia ascertainment is not available.

Corresponding Author: Melinda C. Power, eorge Washington University, Milken Institute School of Public Health, 950 New Hampshire Ave NW, Washington, DC 20052, Tel: 202.994.7778, power@gwu.edu.

Author Contributions: MCP, AC, and KZG all made substantial contributions to the conception and design of the study and analysis and interpretation of the data, MCP drafted the article, AC and KZG revised it critically for important intellectual comment. MCP, AC, and KZG gave final approval of the version to be published.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Statements

Statement of Ethics: Informed consent was provided by HRS and ADAMS participants at data collection. This research was approved by the George Washington University Institutional Review Board.

Reproducibility: The HRS data used in this study are available on the Health and Retirement Study website (http://hrsonline.isr.umich.edu/). ADAMS and HRS-linked Medicare data are available to through standard protocols, detailed on the HRS website. Code for assigning algorithmic diagnoses are available on https://github.com/powerepilab/AD_Algorithm_Development and https://github.com/powerepilab/AD_algorithm_comparison. Dementia probabilities and assignments for the three algorithms appropriate for use in disparities research are available as a user-submitted dataset on the HRS website. Code for the full analyses reported here is available upon request from the study authors.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

Medicare; algorithms; dementia; diagnosis; sensitivity and specificity

Introduction

Observational research on dementia relies on accurate determination of which participants do or do not have dementia. Traditionally, dementia ascertainment for research has relied on consensus diagnoses based on application of formal diagnostic criteria to data collected from in-person, study-based examinations.[1–4] We will refer to this approach as the "criterion standard" diagnosis. However, this approach is time-consuming, costly, and typically cannot be applied retroactively. Thus, researchers are increasingly relying on alternate forms of dementia ascertainment.

Researchers who rely on diagnoses generated during clinical care typically examine electronic medical records or claims databases for specific ICD-9 or ICD-10 codes related to dementia. Participants with those codes are classified as having dementia.[5–7] Algorithmic diagnoses are a second option, and often take one of two forms. In the first, researchers generate a score summarizing performance on cognitive or cognitive and functional assessments[8, 9]; those scoring below the threshold are classified as having cognitive impairment or dementia. In the second, a prediction algorithm is developed using broadly available data to predict criterion-standard diagnoses obtained in a subset of participants. The relevant characteristics from new observations are fed into the algorithm to obtain predicted probabilities. These probabilities are then used to classify participants as having or not having dementia.[9–11]

Compared to the criterion-standard approach, overall sensitivity of alternate dementia ascertainment strategies is typically between 50 and 80%, while specificity is often >90%, [12–16] However, these overall measures mask potentially wide and important differences in measurement error across sub-groups of interest.[15] This can lead to substantial and non-conservative bias, and subsequently to incorrect conclusions, due to differential misclassification.[17] Thus, our objective is to illustrate the potential extent of bias in estimates of prevalence and risk-factor associations resulting from use of algorithmic or Medicare-based approaches as a surrogate for criterion-standard dementia ascertainment in the nationally-representative Health and Retirement Study (HRS).

Materials and Methods

Study sample

We use data from participants of the Health and Retirement Study (HRS) and an HRS sub-study, the Aging, Demographics, and Memory Study (ADAMS). HRS is a U.S. representative longitudinal cohort study of older adults which began in 1992 and which has enrolled new participants approximately every 6 years.[18, 19] Data on sociodemographic and socioeconomic factors, health and health-related behaviors, cognitive testing, and functional status are collected from participants approximately every 2 years. Proxy

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interviews are conducted for participants who are unwilling or unable to respond for themselves. HRS participants aged ≥70 years who completed the 2000 or 2002 HRS questionnaires were sampled for inclusion in ADAMS using a stratified random sampling approach. Ultimately, of the 1,770 sampled, 856 HRS participants were enrolled and completed initial assessment (Wave A, 2001–2003), which included systematic, in-person dementia ascertainment according to standard criteria.[20, 21]

Criterion-Standard Dementia Ascertainment

ADAMS dementia diagnoses were assigned according to DSM-III-R and DSM-IV criteria and confirmed by a consensus expert panel.[20–22] Here, we consider the ADAMS diagnoses to be the criterion-standard diagnoses against which performance of all other dementia ascertainment approaches are compared.

Algorithmic Dementia Ascertainment

We applied five existing algorithms previously created by other research groups (i.e. Herzog-Wallace (H-W)[8], Langa-Kabeto-Weir (L-K-W)[9], Wu[11], Crimmins[9], and Hurd[10]) to ascertain dementia status for each ADAMS/HRS participant at the time of the ADAMS wave A assessment. These algorithms have previously been shown to have substantial variability in sensitivity and specificity across population subgroups.[15] Details of how we applied these algorithms to determine dementia status are available in our previous work.[15] Briefly, for H-W and L-K-W, we compute a score for each participant based on HRS cognitive and functional data, and apply published cut-offs to classify persons as demented or non-demented. Wu, Crimmins, and Hurd, are regression-based algorithms. For Wu and Crimmins we apply published model coefficients to compute predicted probabilities of dementia, and for Hurd we use the predicted probabilities published on the HRS website. Given lack of recommended alternate cut-offs, we classify those with a predicted probability of dementia >50% as having dementia. Details of each algorithm and its current implementation are provided elsewhere,[8–11, 15] and code to reproduce these algorithmic classifications are available on GitHub (https://github.com/powerepilab/AD_algorithm_comparison).

We also classified persons as having dementia according to three new regression-based algorithms we recently developed (i.e. Expert Model, LASSO, and Modified Hurd).[23]Unlike prior algorithms, these algorithms were developed for use in race/ethnic disparities research and have similar out-of-sample sensitivity and specificity across the three major race/ethnicity (non-Hispanic whites, non-Hispanic blacks, and Hispanics).[23] For each algorithm, participants were classified as having dementia if they had a probability of dementia above a specified race-specific cut-off. Details of the development, including rationale for analytical choices, are available elsewhere. Code for reproducing the probabilities and dementia assignments are available on (https://github.com/powerepilab/AD_Algorithm_Development) and dementia probabilities and assignments are also available as a user-submitted dataset on the HRS website.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Medicare-Based Dementia Ascertainment

Using HRS-linked Medicare claims data, we identified persons with dementia diagnosis in the context of received medical care based on the set of ICD-9 codes previously used by Taylor et al. (Supplemental Methods).[12] We classified persons as having dementia if they had an ICD-9 code for dementia in the 3 years prior to the month of the ADAMS Wave A visit, based on prior recommendations.[12]

Covariates

We defined age (years), gender (male/female), education (Less than high school/High school, GED or greater), race/ethnicity (non-Hispanic white/non-Hispanic black), diabetes status (yes/no), independent activities of daily living (IADLs, no difficulty/any endorsement of “some difficulty” with using the phone, managing money, taking medications, shopping for groceries, or preparing hot meals), and respondent status (self-respondent/proxy-respondent) based on data collected at the closest prior HRS assessment (2000 or 2002) to the ADAMS Wave A study visit. All variables were based on self-report except for respondent status, which was documented by HRS staff.

Statistical Methods

We restricted all analyses to non-Hispanic white and non-Hispanic black participants of ADAMS Wave A (n=751) given the small number of persons of other race/ethnicities. For our primary analyses, we further restrict to the 608 participants with non-missing data on all HRS predictors necessary to assign dementia status using the eight algorithms described above (H-W, L-K-W, Wu, Crimmins, Hurd, Expert Model, LASSO, and Modified Hurd). We refer to this sample as the “common sample” throughout. Because Medicare beneficiaries enrolled in HMO plans (i.e. Medicare Advantage) lack complete Medicare claims records, primary analyses considering performance of Medicare- based diagnosis were further restricted to those with at least one month of fee-for-service (FFS) Medicare enrollment during the 3-year lookback period (n=517).

To evaluate the impact of using non-criterion-standard dementia ascertainment on estimates of dementia prevalence, we compared ADAMS estimates of dementia prevalence, overall and by participant characteristics, to corresponding estimates based on each of the other dementia ascertainment approaches. To evaluate the degree of bias due to misclassification introduced into dementia risk factor analyses, we estimated crude and adjusted prevalence ratios quantifying the cross-sectional association between participant characteristics and dementia using Poisson log-linear regression models, and compared those obtained using the ADAMS-based diagnoses to those obtained using algorithmic or Medicare-based diagnoses. We directly compare results using Medicare-based diagnoses in the Medicare sample to the results using the criterion-standard ADAMS diagnoses in the larger common sample, given the assumption frequently invoked in Medicare data analyses that Medicare FFS participants are reasonably representative of the larger U.S. population.

We conducted several sensitivity analyses. We repeated our analyses using all available data for algorithmic dementia classification, resulting in slightly different samples for each approach (range: n=640 to n=751), to understand the impact of small changes in the data on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

algorithm performance. To understand sensitivity related to the Medicare data lookback period, we considered 1-year and 5-year look-back periods. We also re-ran the 3-year lookback Medicare analyses after restricting to those with no HMO enrollment at any point in the 3-year lookback period to understand differences related to requirements for partial versus full Medicare data.

All analyses were conducted using SAS, version 9.4 and RStudio, Version 1.1.423 running R, Version 3.4.3. We report 95% confidence intervals throughout. All estimates are weighted by ADAMS Wave A sampling weights using the *survey* package in R.[24] Informed consent was provided by HRS and ADAMS participants at data collection. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. This research was approved by the George Washington University Institutional Review Board.

Results

Of the 751 eligible ADAMS participants, 608 persons were included in the common sample used for evaluation of algorithmic diagnoses. Of these, 517 had at least 1 month of FFS Medicare participation in the 3-year lookback period. Weighted characteristics were similar across the common sample and Medicare sub-sample (Table 1).

Overall and sub-group specific sensitivities, specificities, and accuracies for each algorithm relative to the ADAMS diagnosis are provided in Supplemental Table 1. Within the common sample, the alternate classification strategies achieved 41–89% sensitivity, 91–100% specificity, and 90–94% accuracy overall in analyses weighted to recover the US age >70 population. Note that the metrics in Supplemental Table 1 differ from previously reported metrics[15, 23] due to differences in the sample across reports; please refer to the previous reports for out-of-sample algorithmic performance.

Overall dementia prevalence estimates varied across dementia ascertainment methods (Figure 1, Supplemental Table 2). Estimates based on Medicare claims and the algorithms created for general use (H-W, L-K-W, Crimmins, Hurd, and Wu) were lower than the ADAMS estimate, reflecting their high specificity and low sensitivity. Prevalence estimates based on algorithms created for use in race/ethnicity disparities research (Expert Model, LASSO, Modified Hurd) were higher than the ADAMS estimate, reflecting their higher sensitivity and lower specificity. Similar patterns persist when considering dementia prevalence among subgroups (Figure 2, Supplemental Table 2). However, the degree to which each approach over- or under-estimated dementia prevalence varied by participant characteristics, reflecting variability in sensitivity and specificity by subgroup (Supplemental Table 1).

The crude estimates of association between participant characteristics and dementia also varied by dementia ascertainment method (Figure 3, Supplemental Table 2). All approaches support an association between older age and increased risk of dementia. However, the estimated magnitude of the association varied, ranging from a prevalence ratio (PR) of 3.0 (95% CI: 1.9, 4.6) for L-K-W to 5.9 (95% CI: 3.3, 10.7) for Hurd, relative to the ADAMS

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

based estimate of 4.0 (95% CI: 2.4, 6.6). For gender, two of the approaches results in a conclusion of approximately double prevalence of dementia in women (H-W PR: 2.2, 95%CI: 1.3, 3.9; Wu PR: 2.0, 95% CI:1.3, 3.1), relative to the more modest and non-significant finding when using the ADAMS dementia diagnoses (PR: 1.3, 95% CI: 0.9, 2.0). Two other algorithms produce point estimates similar to that observed in ADAMS, but unlike ADAMS, suggest a statistically significant difference in the prevalence by gender in this sample (Expert Model PR: 1.5, 95%CI: 1.01, 2.1; LASSO PR: 1.5, 95% CI: 1.1, 2.3). Use of two approaches (H-W and Medicare) would lead to an interpretation about the association between education and dementia that are inconsistent with the ADAMS-based estimate; while the magnitude of the estimate remains similar across H-W (PR: 1.6, 95% CI: 0.9, 2.6) and ADAMS (PR: 1.9, 95% CI: 1.2, 2.9) the Medicare-based estimate is lower (PR: 1.0, 95%CI: 0.6, 1.6). Medicare also performs poorly when estimating the association between diabetes and dementia, yielding a markedly protective association (PR: 0.4, 95% CI 0.2, 0.9) relative to the ADAMS estimate (PR: 0.9, 95% CI: 0.5, 1.7). Estimates of relative prevalence by race across approaches vary in both magnitude and findings for many of the algorithms. The L-K-W and Hurd considerably over-estimate the magnitude of the racial disparity (L-K-W PR: 3.3, 95% CI: 2.2, 5.0 and Hurd PR: 2.7, 95% CI: 1.7, 4.2, versus ADAMS PR: 1.7, 95% CI: 1.1, 2.7), while the Medicare-based estimate would suggest no disparity (PR: 1.2, 95% CI: 0.7, 2.2). Finally, the association between presence of IADLs and dementia is substantially over-estimated using any of the five existing algorithms developed for general use.

The adjusted estimates of association between participant characteristics and dementia also varied by dementia ascertainment method (Figure 4, Supplemental Table 3). The overall pattern of findings for the adjusted associations was similar to that observed in the crude associations for gender, race, and IADLS. Conversely, after adjustment, the H-W algorithm produced an association between age and dementia (PR: 1.0, 95% CI: 0.8, 1.3) that is inconsistent with the ADAMS estimate (PR: 1.9, 95%CI: 1.4, 2.5). Use of Wu now produces an adverse association between education and dementia (PR: 1.3, 95%CI: 1.0, 1.8) while use of H-W and Medicare suggest substantially protective associations (H-W PR: 0.7, 95%CI: 0.5, 0.98; Medicare PR: 0.5, 95%CI: 0.3, 0.8) that are inconsistent with the ADAMS-based estimate (PR: 1.0, 95% CI: 0.7, 1.4). Medicare, and now H-W, suggest protective associations between diabetes and dementia although the 95% confidence intervals contain the null (Medicare PR: 0.5, 95%CI: 0.3, 1.1; H-W PR: 0.6, 95%CI: 0.3, 1.1), contrary to the null estimate derived using the ADAMS-based diagnoses (PR: 1.0, 95%CI: 0.6, 1.6).

Findings of sensitivity analyses using all available data, rather than the common sample, were consistent with primary findings regarding estimates of prevalence overall and within subgroups (Supplemental Table 4), as well as estimates of crude (Supplemental Table 4) and adjusted (Supplemental Table 5) risk factor associations. Sensitivity analyses considering different analytical approaches for Medicare suggest increasing look-back periods yields higher overall and subgroup prevalence estimates, although estimates based on longer look-back periods still under-estimate prevalence relative to the ADAMS-based estimate (Supplemental Table 6). Requiring no HMO Medicare during the 3-year lookback period did not substantially change prevalence estimates (Supplemental Table 6). Findings based on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

measures of association were generally similar regardless of analytical decisions related to use of the Medicare data (Supplemental Table 6, Supplemental Table 7).

Discussion/Conclusion

Our work illustrates that use of algorithmic dementia classification methods or Medicare claims to estimate dementia prevalence will result in findings that differ from those that would be obtained with criterion-standard, in person evaluation. Algorithms maximizing sensitivity over-estimate prevalence (Expert Model, LASSO, and Modified Hurd) while those maximizing specificity under-estimate prevalence (H-W, L-K-W, Crimmins, Hurd, Wu, and Medicare). While this general pattern also holds among subgroups, sensitivity and specificity differences by subgroups often create additional variability in the degree of over- or under-estimation of prevalence by subgroups.

Our work also illustrates that use of algorithms or Medicare claims results in differential misclassification relative to criterion-standard dementia diagnoses that can, but does not always lead to risk-factor associations that substantially differ from those that would be obtained with criterion-standard evaluation. Notably, adjustment for covariates did not eliminate this bias.

These results are generally consistent with recent, related work, which also found overall and race-specific differences in who is identified as having dementia when using, ADAMS, L-K-W and Medicare claims to identify persons with dementia.[25, 26] However, here we took a more comprehensive approach, considering 9 different classification strategies, and also consider how use of different algorithms impacts risk factor associations.

Six of the nine alternate dementia ascertainment methods evaluated (H-W, L-K-W, Crimmins, Wu, Hurd, and Medicare) yield either crude or adjusted measures of association that are substantially inconsistent with the ADAMS estimates in both magnitude and the resulting conclusion about presence/absence or direction of association for at least one considered characteristic. The three remaining approaches (Expert Model, LASSO, and Modified Hurd) generally lead to conclusions about the presence or absence and magnitude of association that are consistent with those we obtained using ADAMS for the example exposures we considered. Therefore, of the methods evaluated, we would recommend use of these three algorithms, although we cannot guarantee that they will have the best performance in all situations. Wherever possible, we recommend comparing performance using the intended algorithm to the criterion standard ADAMS diagnoses before wider application.

This work illustrates the value of validating algorithmic performance in the context of the intended use, including the analysis of interest. Unfortunately, situations in which algorithmic diagnoses are most valuable are also likely to be those where validation is not immediately possible. Instead, we recommend implementing several different, generally well-performing algorithms to evaluate the sensitivity of the results to differences in algorithm performance. Robust findings can later be confirmed in other samples with criterion-standard diagnoses, much as robust case-control or cross-sectional findings can be

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

confirmed with a more expensive cohort study. This work also illustrates the value of algorithm development processes that impose requirements for similar measurement properties among important subgroups. The three algorithms designed to minimize differences in performance across racial/ethnic groups – Expert Model, LASSO, and Modified Hurd – had consistently reasonable performance quantifying associations between risk factors and dementia relative to the other algorithms, which resulted in inconsistent associations for at least one risk factor considered.

Our findings further imply that reports using algorithmic or records-based dementia ascertainment as a surrogate for criterion-standard dementia ascertainment should be interpreted cautiously, particularly where they differ from studies using in-person assessment or when studies using in-person assessment are not available. Dementia ascertainment strategy should be considered in systematic reviews and meta-analyses; where there is evidence of heterogeneity across studies using criterion-standard and algorithmic ascertainment, it may be appropriate to give greater weight to those studies using the criterion-standard approach. In addition, several cohorts have introduced hybrid approaches for dementia ascertainment, leveraging claims or record-based diagnoses for those who do not return for in-person examination.[27, 4] While the desire to maximize follow-up is warranted due to valid concerns about selection bias and loss of power, the utility of this approach must be weighed against the possibility of introducing substantial bias due to differential misclassification, which could result in estimates of association that are either higher or lower than the truth.

Our findings are subject to some caveats. First, our sample size is small, and precision around the point estimates we report is often poor. While confidence intervals around our estimates of crude and adjusted risk-factor associations using algorithmic or Medicare-based diagnoses often overlap with those from the ADAMS estimate, estimates may converge or diverge more obviously in a larger sample. In addition, because our focus is on understanding differences that may be observed using different ascertainment methods, we do not recommend making any substantive conclusions about variation in dementia prevalence by participant characteristics based on the data presented here. Second, the ADAMS data is relatively old; whether algorithmic performance remains constant over time is unclear, and this assumption may be particularly questionable for Medicare-based ascertainment.[26] Third, in the absence of a true “gold standard,” we rely on study-based ADAMS diagnoses as the criterion standard, noting that there may be inaccuracy in ADAMS diagnoses related to participant characteristics. Fourth, as our estimates of bias compare the observed prevalence ratio to a prevalence ratio based on the criterion standard, our estimates of the magnitude of bias actually represent a mix of bias and random variation (isolation of bias would require quantification and comparison of the expected value of the estimate to the truth[28]). The true extent and direction of the bias may differ from what is inferred based on the results from this single sample. Regardless, our results remain useful to illustrate how different an estimate might be relative to the answer that would be obtained using the criterion standard in a given sample. We also report prevalence ratios rather than odds ratios, which are more frequently reported in the literature. As dementia is not rare, differences in the magnitude of point estimates may be more pronounced when using odds ratios. In addition, we focused on associations with prevalent dementia, including persons

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with mild, moderate, and severe dementia. Based on prior work illustrating even greater differences in algorithmic performance when considering incident dementia cases,[15] who typically have less severe symptoms, we expect the potential for bias to be greater when applied to samples enriched in mild dementia cases. Finally, while we believe the overall conclusions are broadly generalizable, performance of individual algorithms may differ in external samples; however, prior work estimating measures of out-of-sample performance suggest overfitting is not a major concern.[15, 23]

We also note that our findings pertain only to the situation where investigators use Medicare-based dementia ascertainment as a surrogate for criterion-based dementia ascertainment. Medicare data can also be used to quantify different, but related quantities that may be of interest depending on the purpose of the study. For example, Medicare claims data can be used to identify whether someone has received a clinical diagnosis of dementia in the medical setting or simply to identify persons with a Medicare claim for dementia. Furthermore, one would expect real differences in prevalence and risk factor associations across these different quantities; these differences are of substantive interest, rather than an indication of bias.

In conclusion, while algorithmic and Medicare-based dementia ascertainment may be useful as a surrogate for criterion-standard dementia ascertainment in some settings, they may result in substantially inaccurate findings in others. No algorithm resulted in accurate estimates of dementia prevalence, although several produced reasonable estimates for risk factor associations. Sub-studies to understand the sub-group specific sensitivity and specificity of algorithmic diagnoses relative to the in-person research-based diagnoses provide valuable insight into when use of algorithmic or Medicare-based diagnoses are likely to be biased. Investigators may wish to limit use of algorithmic or Medicare-based diagnoses to situations where sensitivity and specificity are similar across their exposure or classifier of interest or use multiple algorithms in cases where it is not possible to evaluate algorithm performance differences across levels of exposure. Findings obtained using algorithmic diagnoses should be confirmed in studies using criterion-standard, in-person research-based dementia ascertainment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgments: We thank the study investigators, staff, and participants of the ADAMS and HRS studies.

Funding Sources: This work was funded by R03 AG055485 to MCP from the National Institute on Aging (NIA) at the National Institutes of Health (NIH). The Health and Retirement Study data is sponsored by the National Institute on Aging (NIA/NIH U01AG009740) and was conducted by the University of Michigan. NIA/NIH had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REFERENCES

1. Kukull WA, Higdon R, Bowen JD, et al. Dementia and alzheimer disease incidence: A prospective cohort study. Archives of Neurology. 2002;59(11):1737–46. [PubMed: 12433261]
2. Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and findings from the religious orders study. Curr Alzheimer Res. 2012 7;9(6):628–45. [PubMed: 22471860]
3. Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S. Incidence of Dementia over Three Decades in the Framingham Heart Study. New England Journal of Medicine. 2016 2016/02/11;374(6):523–32. [PubMed: 26863354]
4. Gottesman RF, Albert MS, Alonso A, Coker LH, Coresh J, Davis SM, et al. Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort. JAMA neurology. 2017 8 07.
5. Hung YN, Kadziola Z, Brnabic AJ, Yeh JF, Fuh JL, Hwang JP, et al. The epidemiology and burden of Alzheimer's disease in Taiwan utilizing data from the National Health Insurance Research Database. ClinicoEconomics and outcomes research : CEOR. 2016;8:387–95. [PubMed: 27536149]
6. Goodman RA, Lochner KA, Thambisetty M, Wingo TS, Posner SF, Ling SM. Prevalence of dementia subtypes in United States Medicare fee-for-service beneficiaries, 2011–2013. Alzheimers Dement. 2017 1;13(1):28–37. [PubMed: 27172148]
7. Gilsanz P, Corrada MM, Kawas CH, Mayeda ER, Glymour MM, Quesenberry CP Jr., et al. Incidence of dementia after age 90 in a multiracial cohort. Alzheimers Dement. 2019 Apr;15(4):497–505.
8. Herzog AR, Wallace RB. Measures of cognitive functioning in the AHEAD Study. J Gerontol B Psychol Sci Soc Sci. 1997 5;52 Spec No:37–48. [PubMed: 9215356]
9. 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 7;66 Suppl 1:i162–71. [PubMed: 21743047]
10. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary costs of dementia in the United States. N Engl J Med. 2013 4 4;368(14):1326–34. [PubMed: 23550670]
11. Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Maria Glymour M. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. Alzheimer Dis Assoc Disord. 2013 Jul-Sep;27(3):207–12. [PubMed: 22992720]
12. Taylor DH Jr., Ostbye 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):807–15. [PubMed: 19542620]
13. Duara R, Loewenstein DA, Greig M, Acevedo A, Potter E, Appel J, et al. Reliability and validity of an algorithm for the diagnosis of normal cognition, mild cognitive impairment, and dementia: implications for multicenter research studies. Am J Geriatr Psychiatry. 2010 4;18(4):363–70. [PubMed: 20306566]
14. Katon WJ, Lin EH, Williams LH, Ciechanowski P, Heckbert SR, Ludman E, et al. Comorbid depression is associated with an increased risk of dementia diagnosis in patients with diabetes: a prospective cohort study. J Gen Intern Med. 2010 5;25(5):423–9. [PubMed: 20108126]
15. 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 3;30(2):291–302. [PubMed: 30461528]
16. McGuinness LA, Warren-Gash C, Moorhouse LR, Thomas SL. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review. Pharmacoepidemiol Drug Saf. 2019 2;28(2):244–55. [PubMed: 30667114]
17. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 1977 5;105(5):488–95. [PubMed: 871121]
18. Weir DR, Faul JD, Phillips JWR, Langa KM, Ofstedal MB, Sonnega A. Cohort Profile: the Health and Retirement Study (HRS). International Journal of Epidemiology. 2014;43(2):576–85. [PubMed: 24671021]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

19. Health and Retirement Study. Produced and Distributed by the University of Michigan with Funding from the National Institute on Aging. (Grant Number U01AG009740).
20. Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The Aging, Demographics, and Memory Study: Study Design and Methods. Neuroepidemiology. 2005;25(4):181–91. [PubMed: 16103729]
21. Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. Neuroepidemiology. 2007;29(1–2):125–32. [PubMed: 17975326]
22. Plassman BL, Langa KM, McCammon RJ, Fisher GG, Potter GG, Burke JR, et al. Incidence of dementia and cognitive impairment, not dementia in the United States. Ann Neurol. 2011 9;70(3):418–26. [PubMed: 21425187]
23. 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 1;31(1):126–33. [PubMed: 31567393]
24. Lumley T. Survey: Analysis of complex Survey Samples. R package version 332 2017.
25. 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. [PubMed: 31198838]
26. 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. 2020 6 26.
27. Oudin A, Forsberg B, Adolfsson AN, Lind N, Modig L, Nordin M, et al. Traffic-Related Air Pollution and Dementia Incidence in Northern Sweden: A Longitudinal Study. Environ Health Perspect. 2016 3;124(3):306–12. [PubMed: 26305859]
28. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. Int J Epidemiol. 2005 6;34(3):680–7. [PubMed: 15802377]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

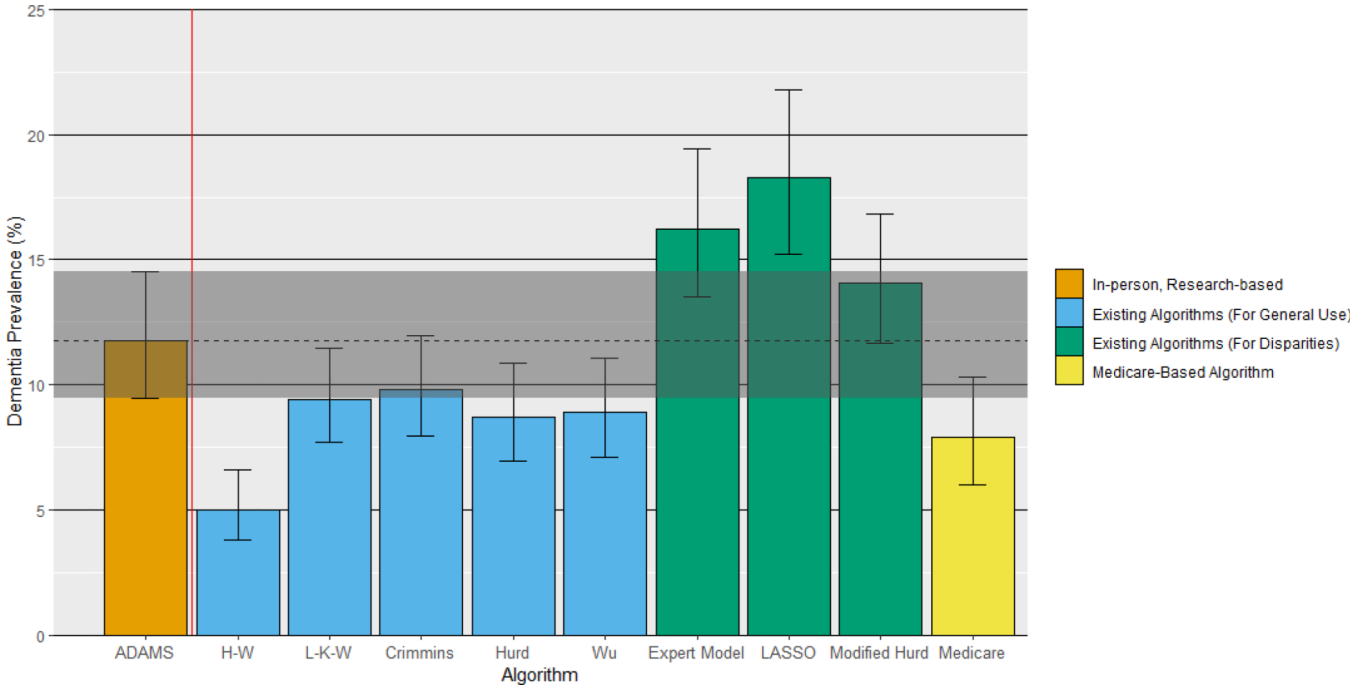

Figure 1.

Comparison of dementia prevalence estimates based on algorithmic and Medicare-based diagnoses to the ADAMS-based estimate

All estimates are based on the common sample (n=608) except for Medicare, which is based on the Medicare sub-sample (n=517), and are weighted using Wave A ADAMS sampling weights. Error bars denote 95% confidence intervals around the dementia prevalence estimate. The shaded band denotes the 95% confidence interval around the ADAMS-based estimate, for comparison.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

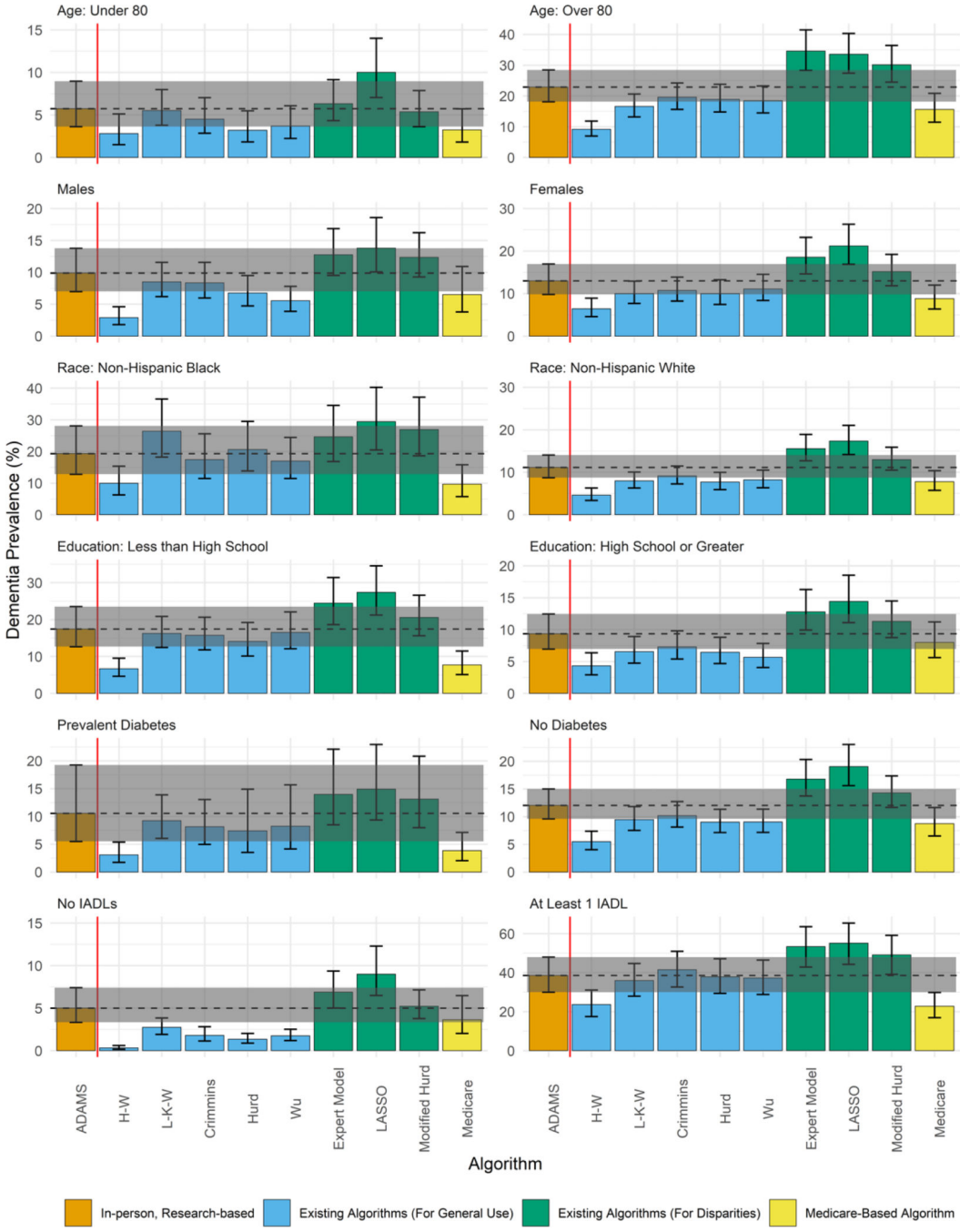

Figure 2.

Comparison of dementia prevalence based on algorithmic and Medicare-based diagnoses to the ADAMS-based estimate by participant characteristics

All estimates are based on the common sample (n=608) except for Medicare, which is based on the Medicare sub-sample (n=517), and are weighted using Wave A ADAMS sampling weights. Error bars denote 95% confidence intervals around the dementia prevalence estimate.

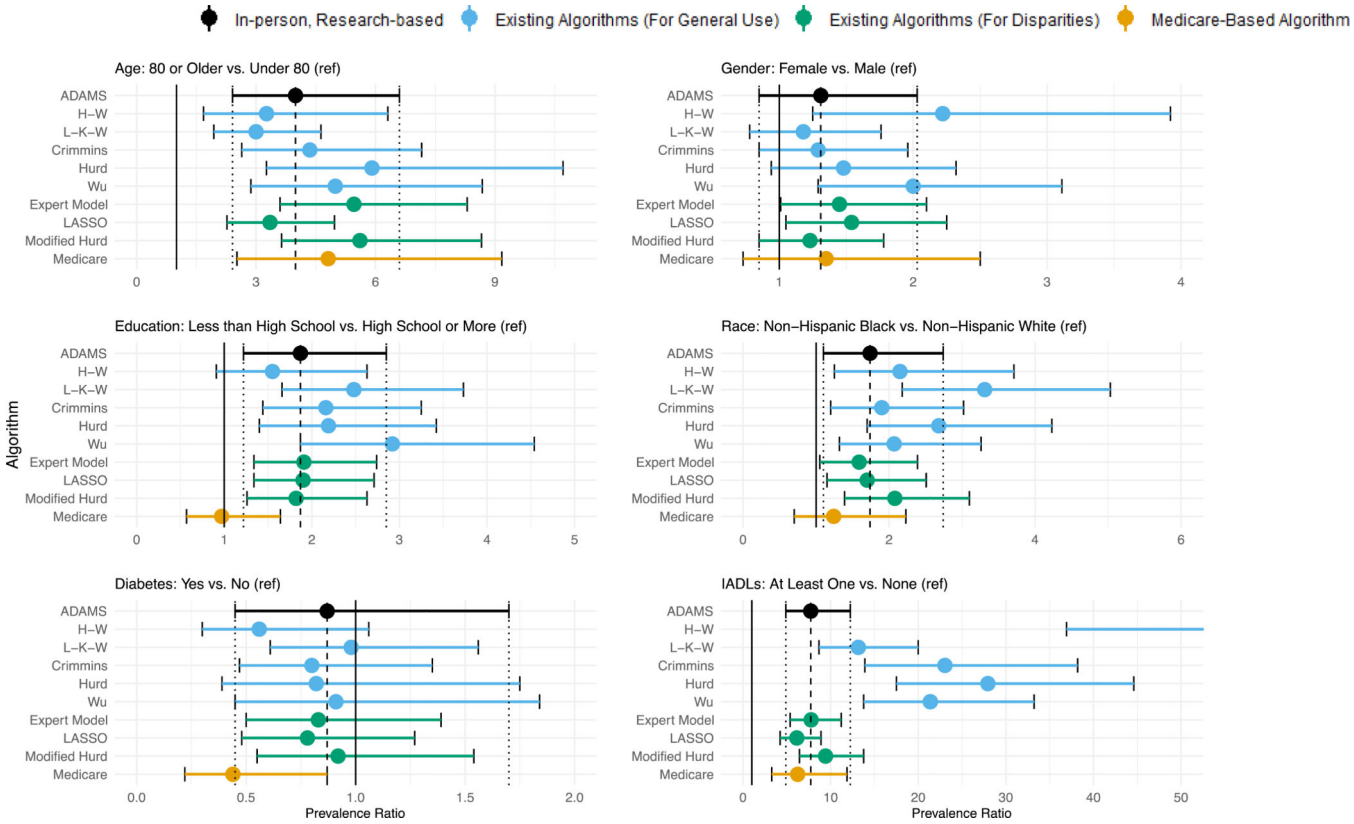

Figure 3.

Comparison of crude associations between participant characteristics and dementia based on algorithmic and Medicare-based diagnoses to the those using the ADAMS-based estimate All estimates are based on the common sample (n=608) except for Medicare, which is based on the Medicare sub-sample (n=517), and are weighted using Wave A ADAMS sampling weights. Error bars denote 95% confidence intervals around the prevalence ratio. The point estimate for the H-W algorithm for the association between IADLs and prevalent dementia is in excess of 50, and so is not shown to avoid compression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

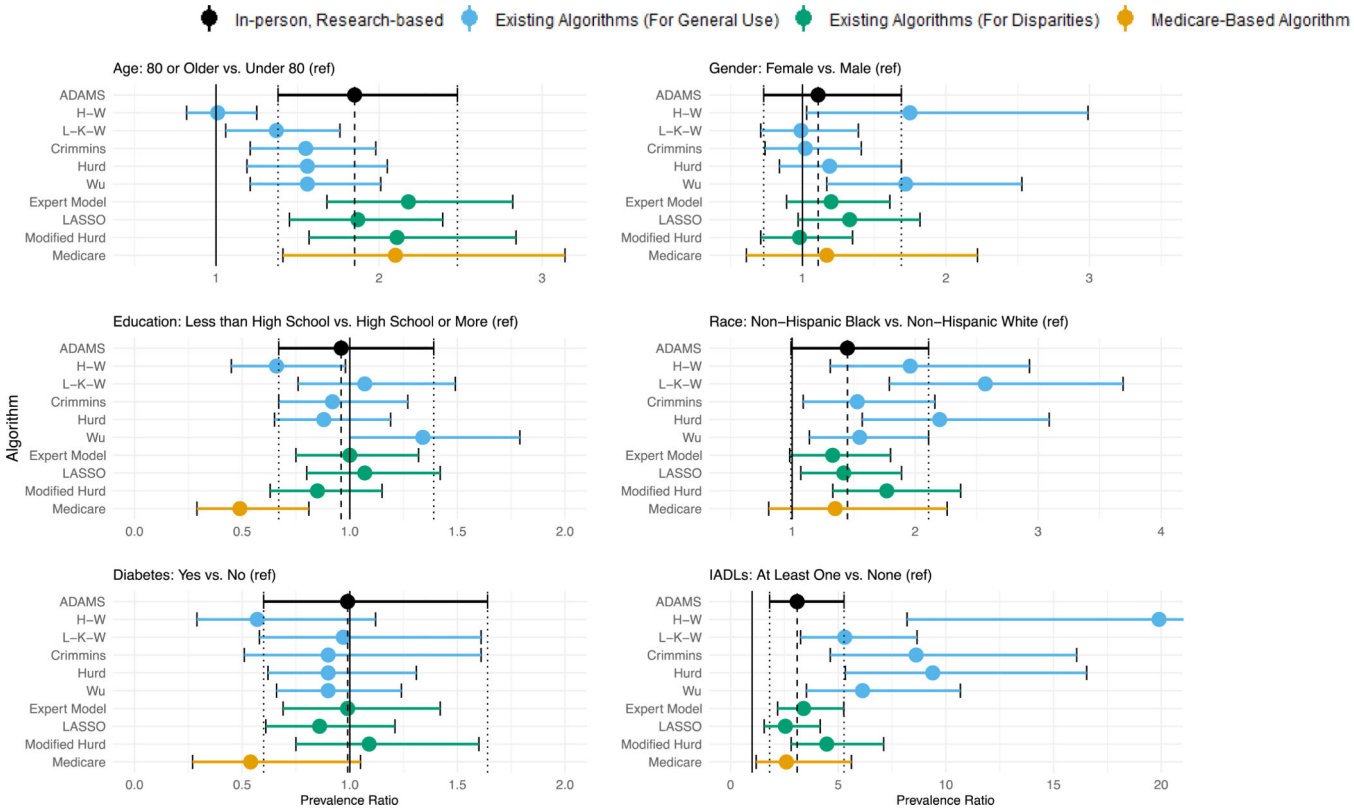

Figure 4.

Comparison of adjusted associations between participant characteristics and dementia based on algorithmic and Medicare-based diagnoses to those using the ADAMS-based estimate

All estimates are based on the common sample (n=608) except for Medicare, which is based on the Medicare sub-sample (n=517), are mutually adjusted for all other characteristics in the figure as well as HRS proxy-respondent status, and are weighted using Wave A ADAMS sampling weights. Error bars denote 95% confidence intervals around the prevalence ratio.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Weighted characteristics of the eligible HRS/ADAMS participants

Characteristic [a]	Common Sample, N=608	Medicare, Any FFS Sub-Sample, N=517
Age in years, mean	77.7	78.1
80 years or older, %	35%	38%
Non-Hispanic Black, %	8%	7%
Female, %	60%	61%
Less than high school, %	30%	30%
Prevalent Diabetes, %	19%	17%
Any IADL limitation, %	20%	22%

[a]All means and % are weighted using Wave A ADAMS sampling weights, which are intended to recover nationally-representative estimates.

Abbreviations: ADAMS, Aging, Demographics, and Memory Study; HRS, Health and Retirement Study; FFS, fee-for-service; IADL, instrumental activities of daily living