How Missing Medication Data Contributes to Bias in Alzheimer's Disease Machine Learning Models

Earlel Thiyagaratnam[1,2], Hawwa Khan[2,3], Apurva Narayan[1]

[1]University of Western Ontario

[2]International Centre for Applied Systems Science and Sustainable Development

[3]McMaster University

ethiyaga@uwo.ca, khanh207@mcmaster.ca, apurva.narayan@uwo.ca

Abstract

Alzheimer's disease (AD) is the most common cause of dementia, yet many cases go undiagnosed due to limited access to expensive brain scans and lab tests. This study investigated whether medication data could help identify AD. Using data from 1,785 participants in the US-representative National Health and Nutrition Examination Survey 2013–2014, we identified 105 individuals (5.9%) with memory test scores suggesting possible AD. We evaluated seven machine learning models using medication features. Models that incorporated contextual prescription information, including the reasons for medication use and conditions being treated, achieved the best performance (AUC 0.61–0.63). In contrast, models using only basic drug names or provider information performed poorly (AUC 0.46–0.51). This performance difference was statistically significant ($t = 14.98$, $p < 0.0001$). Our findings suggest that medications data, when analyzed with attention to clinical context, could serve as a low-cost tool for identifying individuals at risk of AD. This approach may help address diagnostic disparities in settings with limited access to advanced testing.

Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder and the leading cause of dementia, faces diagnostic inequity (Alzheimer's Association 2024). Advanced tools like neuroimaging and biomarkers are often inaccessible in low-resource settings due to cost and infrastructure needs. To bridge this gap, we evaluate low-cost, accessible variables for predicting AD. Moving beyond isolated factors, we use a systems approach to understand how individual and environmental variables interact within health and social contexts, uncovering potential diagnostic biases. Leveraging the US-representative National Health and Nutrition Examination Survey (NHANES) 2013–2014 medications dataset, and a (Consortium to Establish a Registry for Alzheimer's Disease) CERAD-based cognitive function module, we address two questions: (1) How does variable exclusion introduce bias in AD detection models? (2) Can contextual variables effectively substitute for unavailable biomarkers?

Background

AD stems from complex interactions among biological, environmental, cultural, and psychological factors. Medications play a complicated role, with some potentially slowing the disease and others having unclear or harmful effects on cognition. For example, AD-specific drugs like lecanemab, which targets amyloid-beta, have shown promise in slowing cognitive decline in early-stage AD (van Dyck et al. 2023). Meanwhile, medications for heart and metabolic conditions, such as ezetimibe and statins, may affect the brain, though their direct impact on AD remains uncertain (Nußbaumer et al. 2016). Anti-inflammatory and cholesterol lowering drugs also require careful consideration, as some may reduce brain inflammation, while others might worsen cognitive issues (Xu et al. 2022). Therefore, doctors should prescribe medications for adults at risk of or living with AD based on both their primary purpose and potential long-term effects on the brain. Wise medication management, guided by current research, is essential to support cognitive health.

The NHANES medication dataset contains important information that can help us understand how drugs may affect AD. For example, the variable RXDUSE indicates whether, in the past 30 days, the respondent has used or taken a prescription medication (excluding prescription vitamins or minerals already reported). Older adults often take multiple medications, and this can affect cognitive health, either positively or negatively, depending on the type of drug (Gnjidic et al. 2012; Livingston et al. 2020). The variables RXDDRUG and RXDDRGID list the generic drug name and code, which helps us study how certain drugs, like cholinesterase inhibitors or anticholinergics, may impact memory and dementia risk (Richardson et al. 2018). RXQSEEN indicates whether a prescription container was seen by interviewer, which is important because proper medical oversight can prevent harmful drug effects and cognitive decline (Fick et al. 2019).

RXDDAYS records how many days the respondent has been using or taking drug, helping us study how consistent use affects brain health. For example, regular use of heart or diabetes medications may protect against cognitive decline, while irregular use may not (Geifman et al. 2017). RXDRSC1/2/3 list the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnosis codes linked to the reported prescription drug, starting from the first diagnosis to third diagnosis listed, which is useful because drugs for conditions like high blood pressure or depression may indirectly influence AD risk (Livingston et al. 2017). Similarly, RXDRSD1/2/3 note the ICD-10-CM code descriptions, and since chronic illnesses like heart disease and inflammation are linked to cognitive decline, managing them properly may affect AD risk (Iturria-Medina et al. 2014; Xu et al. 2022). RXDCOUNT note the number of prescription medicines reported. Taking many drugs at once can increase the risk of harmful drug interactions and side effects, which may worsen cognitive function (Gnjidic et al. 2012).

By studying NHANES medication data, we can better understand how different medication variables interact with one another and its effect on AD. This helps us understand how impactful medication data is for AD prediction.

Methods

Using the 2013–2014 NHANES cognitive module, we created a binary indicator of possible AD-related impairment from the CERAD Word List Memory Test (immediate recall, three trials; delayed recall). Participants were flagged if delayed recall ≤ 2 and immediate recall < 4, thresholds grounded in diagnostic research and National Institute on Aging and Alzheimer's Association (NIA-AA) guidelines. Scores that fall about 1.5 to 2 standard deviations (SDs) below what is normal for someone's age can help identify memory problems such as difficulty learning and remembering new information, which are common in AD (Welsh et al. 1992; Petersen et al. 1999; Chandler et al. 2005; Seo et al. 2010), improving specificity over using one measure approaches. Using this method, 105 individuals (5.9% of the sample) were identified. The variable reflects AD patterns rather than formal diagnoses and requires clinical validation.

The 2013–2014 NHANES medication dataset was then merged with the AD indicator using the participant identifier (SEQN). This dataset contains structured, coded variables such as drug identifiers (RXDDRUG, RXDDRGID), therapeutic class codes (RXDRSC1/2/3), subclass codes (RXDRSD1/2/3), and numeric indicators of use and duration (RXDUSE, RXQSEEN, RXDDAYS, RXDCOUNT). Because NHANES provides these variables in standardized formats rather than as free-text drug descriptions, no additional processing was required. Instead, the preprocessing step involved extracting all relevant numeric variables, excluding identifiers and outcome variables, and applying mean imputation to replace missing values rather than discarding incomplete records.

All one- to five-feature combinations were tested to balance predictive power, interpretability, and computational cost. This range, consistent with the "one in ten rule" (Peduzzi et al. 1996), mitigates overfitting in small datasets while allowing detection of synergistic effects between features (Saeys, Inza, and Larrañaga 2007; Raschka 2020; Pudjihartono et al. 2022). We tested seven different machine learning (ML) models: Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB), because each works differently and can capture different kinds of patterns in the data. To check how well they would work on new data, we used three-fold cross-validation, which splits the data into three parts, two parts for training and one part for testing (Demšar 2006; Wong and Yeh 2020).

For each set of features, we measured model performance using the area under the curve (AUC). AUC is useful because it measures how well a model can separate people with and without AD, and it lets us compare models on the same scale (Labatut and Cherifi 2011; Brown 2018). Then, we kept the top five and bottom five sets for more detailed comparisons. To see if the models' performance differences were meaningful rather than due to random chance, we used Tukey's Honestly Significant Difference (HSD) test, which is designed for comparing many groups at once without increasing the chance of false positives (Nanda et al. 2021). Within the top and bottom groups, we used pairwise t-tests with Bonferroni correction to keep the error rate low when making many comparisons. We also ran an independent t-test to see if the top and bottom groups were different.

Finally, we created a heatmap to show the AUC results for each model and feature set visually, making it easy to compare patterns at a glance. To support transparency and reproducibility, all preprocessing scripts and codes used for analysis, along with references to the publicly available NHANES datasets, will be shared in a GitHub repository.

Findings

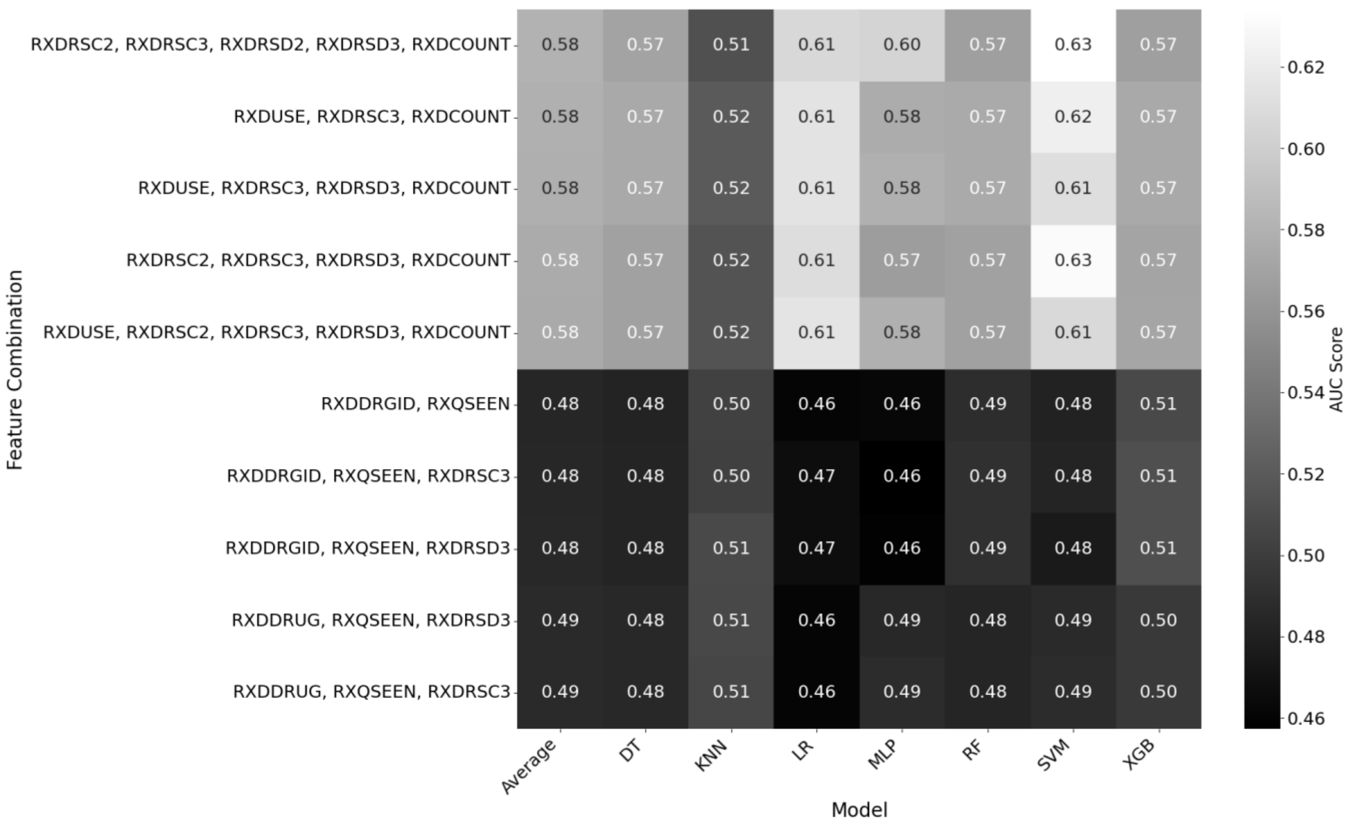

Figure 1: NHANES 2013–2014 medications heatmap.

The top feature combination is a combination of second and third diagnosis ICD-10-CM code/description for drug use and total drug count ({RXDRSC2/3, RXDRSD2/3, and RXDCOUNT}).

In figure 1, our ML model performance showed limited AUC variation, with scores ranging from 0.46 to 0.63. These differences were driven more by feature construction than model choice. The top five feature sets yielded moderate predictive accuracy (AUC 0.57–0.63), while the bottom five performed poorly (AUC 0.46–0.51), a 0.15 difference. SVM and LR consistently outperformed other models, with SVM achieving the highest AUC using features related to RXDRSC2/3, RXDRSD2/3, and RXDCOUNT. In contrast, KNN and XGB were less reliable. Features like RXDDRGID and RXQSEEN offered limited predictive value. This suggests that combining clinically relevant medication features with models like SVM or LR yields the best results, though additional data types are needed to reach clinically actionable performance.

In table 1, statistical testing confirmed these performance differences. Tukey's HSD test revealed that all pairwise model comparisons were statistically significant ($p < 0.05$), including small differences such as between RF and XGB ($\Delta = 0.0036$, $p = 0.0066$), indicating that even modest AUC gaps were robust. LR and SVM significantly outperformed DT, reinforcing the importance of model selection.

Summary Metric	Result
Total model pair comparisons (DT, KNN, LR, MLP, RF, SVM, XGB)	21
Significant differences (p-adj ≤ 0.05)	21 (100%)
Largest positive mean difference	0.0459 (SVM vs. XGB)
Largest negative mean difference	-0.0650 (LR vs. SVM)
Smallest absolute mean difference	0.0036 (RF vs. XGB)

Table 1: Summary P-values of ML models (Tukey's HSD).

Test	Value	Significant
T-statistic	14.9813	True
P-value	0.0000	True

Table 2: Independent t-test comparing top 5 and bottom 5 feature sets.

Among the top five feature sets, pairwise t-tests with Bonferroni correction found no significant differences (all p > 0.05), even when comparing simpler and more complex combinations. For example, in table 1.A, a three-feature set ({RXDRSC3, RXDUSE, and RXDCOUNT }) performed comparably to a five-feature set ({RXDRSC2/3, RXDRSD2/3, and RXDCOUNT}; p = 0.9623). Once a core set of informative features is included, performance stabilizes, allowing flexibility in feature selection. The bottom five feature sets also showed no significant differences (all p > 0.87), pointing to uniformly poor predictive value and indicating a floor effect: these features likely lack meaningful association with AD risk (Table 2.A).

In table 2, an independent samples t-test comparing top and bottom feature groups confirmed a significant performance gap (t = 14.9813, p < 0.0001), with minimal variability within each group but a large effect size between them.

Discussion

NHANES Medication Data: Construct Validity for AD

In their paper, Ligang Liu et al. (2024) used NHANES reasons-for-use data to study usual and off-label medication use in the population, showing these variables are reliable for understanding clinical context and treatment intent. Their work used NHANES data to assess whether medications matched their intended use, supporting the idea that RXDRSC2/3 and RXDRSD2/3 contain meaningful clinical information rather than random noise.

These variables are relevant to AD risk because RXDRSC2/3 and RXDRSD2/3 often include conditions like diabetes, high cholesterol, and depression. These conditions are known modifiable risk factors for AD, which focus on vascular and mental health pathways to cognitive decline (Livingston et al. 2017; Livingston et al. 2020). Therefore, RXDRSC2/3 and RXDRSD2/3 provide insight into comorbidity patterns linked to AD risk, not just medication use.

RXDCOUNT adds another layer by measuring polypharmacy, which has been linked to cognitive decline and AD. Possible reasons include anticholinergic drug effects, sedative use, drug interactions, and treatment burden leading to delirium or faster decline (Gnjidic et al. 2012; Yu et al. 2024; Salahudeen, Hilmer, and Nishtala 2015). Anticholinergic drugs, in particular, have shown a dose response relationship with AD risk in well-controlled studies (Coupland et al. 2019; Richardson et al. 2018). While NHANES does not calculate anticholinergic burden directly, RXDCOUNT and the reasons/conditions fields together help estimate two key medication-related risks: total drug load and the conditions for which high-risk medications are prescribed.

Finally, the outcome measure uses the CERAD word list tests from NHANES 2011–2014, which assess memory encoding and delayed recall. These tests capture memory problems typical of early AD, making them relevant for studying medication effects. In her paper, Debra Brody (2019) provides validation for the CERAD Word List in NHANES. Using delayed recall with an immediate-recall cutoff aligns with prior research on AD (Brody 2019).

Convergent Validity in Medication Data and AD

Our results show that combining RXDRSC2/3 with RXDCOUNT works better than just RXDDRUG and RXDDRGID. This matches earlier research showing that measures which consider the context of medication use, such as the reason for prescription or overall medication burden, have stronger ties to cognitive outcomes than simple drug counts. For instance, studies have found that long-term use of anticholinergic drugs at higher doses over many years increases the risk of AD (Coupland et al. 2019; Richardson et al. 2018). Other research shows that polypharmacy is linked higher AD risk, even after accounting for health and social factors (Yu et al. 2024). This supports our method, which includes both the purpose of medications and the total medication burden.

Our approach aligns with well-established benchmarks like the Drug Burden Index (DBI; Hilmer et al. 2007) and the Charlson Comorbidity Index (CCI; Charlson et al. 1987). These indices reliably measure medication burden and predict negative cognitive and functional outcomes. Studies have also shown clear correlations between these validated measures and cognitive tests, supporting their accuracy (Gwarieh and Shaker 2003). While our study does not directly compare our combined medication features with these indices, the strong predictive performance and clinical relevance of our features suggest they measure a similar underlying concept. Future research should include direct comparisons and correlation analyses with established indices like the DBI to confirm this connection.

External Validity of Medication Data for AD

The NHANES is a national survey that represents the U.S. population, so the medication data reflect both how diseases are treated and how doctors prescribe medications to older adults in different communities. This makes the results more useful than studies from single hospitals or health systems, which may not capture the full picture of real-world care. Past research has confirmed that NHANES correctly records why people take certain medications (Liu et al. 2024). While NHANES does not follow people over time, it uses CERAD, which connect to broader signs of AD (Brody 2019). Our study also looks at other health problems, which matters because conditions like heart disease can affect AD risk (Livingston et al., 2020).

We also need to consider whether these findings will stay relevant in the future. In their study, Shadish, Cook, and Campbell (2002) explain changes in healthcare or population health over time can make it hard to apply old data to new situations. However, because key factors like the reason for taking a medication and the number of drugs used kept predicting AD risk across different models, these patterns may hold up over time (Munger 2023).

Finally, our study's practical use is important. Models like LR and SVM worked best when using simple, health-related features, which is meaningful in practice. Futoma et al. (2020) say that models should not just be statistically accurate but also help in practice to make decisions. Similarly, in their study, Wan, Caffo, and Vedula (2022) explain that for models to be used in practice, they need to be stable, easy to understand, and relevant to patient care. The fact that the best-performing features in our study stayed consistent suggests that even models with moderate accuracy could still help detect possible AD cases, though more testing in real patient data is still needed.

Comparison to Gold-Standard Methods

The modeling approach uses seven standard classifiers with three-fold cross-validation, measures performance using AUC, and applies multiple-comparison adjustments. This follows common practices in predictive epidemiology and clinical machine learning, particularly for medium-sized health datasets like NHANES (Collins et al. 2015; Steyerberg 2019). We also focused on creating features based on medical expertise, which often matters more than the specific algorithm choice when working with smaller datasets (Allam et al. 2019; Riley et al. 2020).

In similar studies on AD risk using health records, the most common baseline models are logistic regression, Cox proportional hazards models, RF, XGB, and SVM. Many large studies find that after careful feature engineering and calibration, ML models often perform similarly to regularized logistic regression, especially when predictors have mostly linear or consistent effects (Allam et al. 2019). Our results, where SVM and LR performed best while RF were competitive but not superior, match these findings.

K-fold cross-validation is a standard method for internal validation. Research guidelines recommend repeated or nested cross-validation to reduce over-optimism and keep model selection separate from performance evaluation. Our use of cross-validated AUC, multiple-comparison adjustments, and independent testing follows these recommendations. In the future we could involve nested cross-validation to fully separate feature selection from performance estimation (Collins et al. 2015; Steyerberg 2019).

Many established AD risk models in claims or electronic health record studies use diagnosed dementia cases as outcomes with time-to-event models. In contrast, population surveys like NHANES rely on cognitive test scores. The NHANES cognitive module is well studied and widely used in research on cognitive impairment (Brody 2019; Qian, Liu, and Li 2025). This supports our use of a CERAD-based measure as a reasonable outcome for model development and variable selection.

Studies examining medication effects on AD usually categorize drugs by type or count total prescriptions rather than listing individual drug names. Our approach, which combines medical conditions with prescription counts, fits this standard practice. It also aligns with AD medication research that focuses on risk linked to specific drug uses and overall exposure levels (Coupland et al. 2019; Richardson et al. 2018; Yu et al. 2024).

Conclusion

Our study demonstrates that variable selection and design are as crucial as ML methods for AD prediction. Analyzing NHANES, we found medication details (usage reasons, treated conditions, and total drug counts) outperformed simple prescription records in predicting risk. While SVM and LR performed best, variable quality impacted results more than model choice. Our findings show that excluding clinical context might reduce accuracy and introduce bias.

However, these findings should be interpreted with caution. Predictive accuracy remained modest (AUC 0.61–0.63), indicating that medication-based models only are better suited as screening aids than diagnostic tools. Reliance on CERAD scores as a proxy for AD, the small number of impaired cases, and dataset imbalance limit generalizability. This approach enables fairer, low-cost diagnostics where advanced testing is unavailable, though further refinement is needed for clinical use. Benchmarking against indices such as the DBI or CCI, as well as testing across subgroups, could provide stronger clinical models.

Future work should incorporate longitudinal data, additional health and social factors, and validation across di-verse populations to develop equitable screening tools. Future research should also integrate medication data with other accessible measures, such as demographics, lifestyle factors, or brief cognitive tests, and apply longitudinal and external validation to strengthen both accuracy and equity in AD risk prediction.

Appendix

Groups	RXDDRGID, RXQSEEN	RXDDRUG, RXQSEEN, RXDRSC3	RXDDRUG, RXQSEEN, RXDRSD3	RXDDRGID, RXQSEEN, RXDRSC3	RXDDRGID, RXQSEEN, RXDRSD3
RXDDRGID, RXQSEEN	—	0.7100	0.7440	0.9060	0.8927
RXDDRUG, RXQSEEN, RXDRSC3	0.7100	—	0.9610	0.8183	0.8410
RXDDRUG, RXQSEEN, RXDRSD3	0.7440	0.9610	—	0.8518	0.8734
RXDDRGID, RXQSEEN, RXDRSC3	0.9060	0.8183	0.8518	—	0.9846
RXDDRGID, RXQSEEN, RXDRSD3	0.8927	0.8410	0.8734	0.9846	—

Table 1.A: P-values between top 5 feature combinations (Bonferroni-corrected t-tests).

Groups	RXDUSE, RXDRSC3, RXDCOUNT	RXDUSE, RXDRSC3, RXDRSD3, RXDCOUNT	RXDRSC2, RXDRSC3, RXDRSD3, RXDCOUNT	RXDUSE, RXDRSC2, RXDRSC3, RXDRSD3, RXDCOUNT	RXDRSC2, RXDRSC3, RXDRSD2, RXDRSD3, RXDCOUNT
RXDUSE, RXDRSC3, RXDCOUNT	—	0.9465	0.8509	0.8081	0.9623
RXDUSE, RXDRSC3, RXDRSD3, RXDCOUNT	0.9465	—	0.8965	0.8546	0.9125
RXDRSC2, RXDRSC3, RXDRSD3, RXDCOUNT	0.8509	0.8965	—	0.9679	0.8265
RXDUSE, RXDRSC2, RXDRSC3, RXDRSD3, RXDCOUNT	0.8081	0.8546	0.9679	—	0.7868
RXDRSC2, RXDRSC3, RXDRSD2, RXDRSD3, RXDCOUNT	0.9623	0.9125	0.8265	0.7868	—

Table 2.A: P-values between bottom 5 feature combinations (Bonferroni-corrected t-tests).

References

Alzheimer's Association. 2024. 2024 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 20(5): 3708–3821. doi.org/10.1002/alz.13809.

Brody, D. J. 2019. Cognitive performance in adults aged 60 and over: Vol. no. 126. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Brown, J. B. 2018. Classifiers and their metrics quantified. *Molecular Informatics*, 37(1700127): 1–11.

doi.org/10.1002/minf.201700127.

Collins, G. S.; Reitsma, J. B.; Altman, D. G.; and Moons, K. G. M. 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *British Medical Journal*, 350(jan07 4): 1–9.

doi.org/10.1136/bmj.g7594.

Charlson, M. E.; Pompei, P.; Ales, K. L.; and MacKenzie, C. R. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases* 40(5): 373–383. doi.org/10.1016/0021-9681(87)90171-8

Coupland, C. A. C.; Hill, T.; Dening, T.; Morriss, R.; Moore, M.; and Hippisley-Cox, J. 2019. Anticholinergic drug exposure and the risk of dementia: a nested case-control study. J*ournal of the American Medical Association Internal Medicine*, 179(8): 1084–1093. doi.org/10.1001/jamainternmed.2019.0677.

Fick, D. M.; Semla, T. P.; Steinman, M.; Beizer, J.; Brandt, N.; Dombrowski, R.; DuBeau, C. E.; Pezzullo, L.; Epplin, J. J.; Flanagan, N.; Morden, E.; Hanlon, J.; Hollmann, P.; Laird, R.; Linnebur, S.; and Sandhu, S. 2019. American Geriatrics Society 2019 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. *Journal of the American Geriatrics Society*, 67(4): 674–694. doi.org/10.1111/jgs.15767.

Futoma, J.; Simons, M.; Panch, T.; Doshi-Velez, F.; and Celi, L. A. 2020. The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health* 2(9): 489–492. doi.org/10.1016/S2589-7500(20)30186-2.

Geifman, N.; Brinton, R. D.; Kennedy, R. E.; Schneider, L. S.; and Butte, A. J. 2017. Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease. *Alzheimer's Research & Therapy*, 9(1): 1–10. doi.org/10.1186/s13195-017-0237-y.

Gnjidic, D.; Hilmer, S. N.; Blyth, F. M.; Naganathan, V.; Waite, L.; Seibel, M. J.; McLachlan, A. J.; Cumming, R. G.; Handelsman, D. J.; and Le Couteur, D. G. 2012. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. *Journal of Clinical Epidemiology*, 65(9): 989–995.

doi.org/10.1016/j.jclinepi.2012.02.018.

Hilmer, S. N.; Mager, D. E.; Simonsick, E. M.; Cao, Y.; Ling, S. M.; Windham, B. G.; Harris, T. B.; Hanlon, J. T.; Rubin, S. M.; Shorr, R. I.; Bauer, D. C.; and Abernethy, D. R. 2007. A drug burden index to define the functional burden of medications in older people. *Archives of Internal Medicine* 167(8): 781–787.

doi.org/10.1001/archinte.167.8.781.

Iturria-Medina, Y.; Sotero, R. C.; Toussaint, P. J.; and Evans, A. C. 2014. Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. *Public Library of Science Computational Biology*, 10(11): 1–16. doi.org/10.1371/journal.pcbi.1003956.

Labatut, V.; and Cherifi, H. 2011. Evaluation of performance measures for classifiers comparison. arXiv preprint. arXiv:1112.4133 [cs.LG]. Ithaca, NY: Cornell University Li- brary.

Livingston, G.; Huntley, J.; Sommerlad, A.; Ames, D.; Ballard, C. G.; Banerjee, S.; Brayne, C.; Burns, A.; Cohen-Mansfield, J.; Cooper, C.; Costafreda, S. G.; Dias, A.; Fox, N.; Gitlin, L. N.; Howard, R.; Kales, H. C.; Kivimaki, M.; Larson, E. B.; Ogunniyi, A.; … Mukadam, N. 2020. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248): 413–446. doi.org/10.1016/S0140-6736(20)30367-6.

Livingston, G.; Sommerlad, A.; Orgeta, V.; Costafreda, S. G.; Huntley, J.; Ames, D.; Ballard, C.; Banerjee, S.; Burns, A.; Cohen-Mansfield, J.; Cooper, C.; Fox, N.; Gitlin, L. N.; Howard, R.; Kales, H. C.; Larson, E. B.; Ritchie, K.; Rockwood, K.; Sampson, E. L.; … Mukadam, N. 2017. Dementia prevention, intervention, and care. *The Lancet*, 390(10113): 2673–2734. doi.org/10.1016/S0140-6736(17)31363-6.

Liu, L.; Tao, H.; Xu, J.; Liu, L.; and Nahata, M. C. 2024. Quantity, duration, adherence, and reasons for dietary supplement use among adults: results from NHANES 2011–2018. *Nutrients*, 16(1830): 1–15. doi.org/10.3390/nu16121830.

Munger, K. 2023. Temporal validity as meta-science. *Research & Politics* 10(3). 1–10. doi.org/10.1177/205316802311872.

Nanda, A.; Mohapatra, B. B.; Mahapatra, A. P. K.; and Mahapatra, A. P. K. 2021. Multiple comparison test by Tukey's honestly significant difference (HSD): do the confidence level control type I error. *International Journal of Statistics and Applied Mathematics*, 6(1): 59–65. doi.org/10.22271/maths.2021.v6.i1a.636.

Nußbaumer, B.; Glechner, A.; Kaminski-Hartenthaler, A.; Mahlknecht, P.; and Gartlehner, G. 2016. Ezetimibe-statin combination therapy: efficacy and safety as compared with statin monotherapy — a systematic review. *Deutsches Ärzteblatt International*, 113(26): 445–453. doi.org/10.3238/arztebl.2016.0445.

Petersen, R. C.; Smith, G. E.; Waring, S. C.; Ivnik, R. J.; Tangalos, E. G.; and Kokmen, E. 1999. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology*, 56(3): 303–308. doi.org/10.1001/archneur.56.3.303.

Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T. R.; and Feinstein, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12): 1373–1379. doi.org/10.1016/S0895-4356(96)00236-3.

Pudjihartono, N.; Fadason, T.; Kempa-Liehr, A. W.; and O'Sullivan, J. M. 2022. A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, 2(2022): 1–17. doi.org/10.3389/fbinf.2022.927312.

Qian, Y.; Liu, Q.; and Li, T. 2025. Association between composite dietary antioxidant index and cognitive function impairment in the elderly: evidence from NHANES 2011–2014. *Frontiers in Neurology*, 16: 1–12. doi.org/10.3389/fneur.2025.1529989.

Raschka, S. 2020. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint. arXiv:1811.12808 [cs.LG]. Ithaca, NY: Cornell University Li- brary.

Richardson, K.; Fox, C.; Maidment, I.; Steel, N.; Loke, Y. K.; Arthur, A.; Myint, P. K.; Grossi, C. M.; Mattishent, K.; Bennett, K.; Campbell, N. L.; Boustani, M.; Robinson, L.; Brayne, C.; Matthews, F. E.; and Savva, G. M. 2018. Anticholinergic drugs and risk of dementia: case-control study. *British Medical Journal*, 361(1315): 1–12. doi.org/10.1136/bmj.k1315.

Riley, R. D.; Ensor, J.; Snell, K. I. E.; Harrell, F. E.; Martin, G. P.; Reitsma, J. B.; Moons, K. G. M.; Collins, G.; and van Smeden, M. 2020. Calculating the sample size required for developing a clinical prediction model. *British Medical Journal*, 368(441): 1–12. doi.org/10.1136/bmj.m441.

Saeys, Y.; Inza, I.; and Larrañaga, P. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19): 2507–2517. doi.org/10.1093/bioinformatics/btm344.

Salahudeen, M. S.; Hilmer, S. N.; and Nishtala, P. S. 2015. Comparison of anticholinergic risk scales and associations with adverse health outcomes in older people. *Journal of the American Geriatrics Society*, 63(1): 85–90. doi.org/10.1111/jgs.13206.

Shadish, W. R.; Cook, T. D.; and Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Steyerberg, E. W. 2019. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd edition. Springer Nature. doi.org/10.1007/978-3-030-16399-0.

van Dyck, C. H.; Swanson, C. J.; Aisen, P.; Bateman, R. J.; Chen, C.; Gee, M.; Kanekiyo, M.; Li, D.; Reyderman, L.; Cohen, S.; Froelich, L.; Katayama, S.; Sabbagh, M.; Vellas, B.; Watson, D.; Dhadda, S.; Irizarry, M.; Kramer, L. D.; and Iwatsubo, T. 2023. Lecanemab in early Alzheimer's disease. *The New England Journal of Medicine*, 388(1): 9–21. doi.org/10.1056/NEJMoa2212948.

Wan, B., Caffo, B., & Vedula, S. S. (2022). A Unified Framework on Generalizability of Clinical Prediction Models. *Frontiers in Artificial Intelligence*, *5* (872720) 303–308. doi.org/10.3389/frai.2022.872720.

Welsh, K. A.; Butters, N.; Hughes, J. P.; Mohs, R. C.; and Heyman, A. 1992. Detection and staging of dementia in Alzheimer's disease: use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology*, 49(5): 448–452. doi.org/10.1001/archneur.1992.00530290030008.

Wong, T.-T.; and Yeh, P.-Y. 2020. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8): 1586–1594.

doi.org/10.1109/TKDE.2019.2912815.

Xu, J.; Ma, C.; Hua, M.; Li, J.; Xiang, Z.; and Wu, J. 2022. CNS and CNS diseases in relation to their immune system. *Frontiers in Immunology*, 13(1063928): 1–12.

doi.org/10.3389/fimmu.2022.1063928.

Yu, X.; Qian, Y.; Zhang, Y.; Chen, Y.; and Wang, M. 2024. Association between polypharmacy and cognitive impairment in older adults: a systematic review and meta-analysis. *Geriatric Nursing*, 59(2024): 330–337. doi.org/10.1016/j.gerinurse.2024.07.005.