DOI: 10.1002/alz.066538

Alzheimer's & Dementia
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

PUBLIC HEALTH

POSTER PRESENTATION

Comparison of Algorithms for Dementia Classification in the Survey of Health, Ageing and Retirement in Europe

Matthias Klee[1] | Kenneth M Langa[2,3] | Anja K Leist[1]

[1]University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]University of Michigan, Ann Arbor, MI, USA

[3]Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA

Correspondence
Matthias Klee, University of Luxembourg, Esch-sur-Alzette, Luxembourg.
Email: matthias.klee@uni.lu

Abstract

Background: Low dementia rates, reflecting underdiagnosis in representative cohort studies, limit statistical power of etiological and preventative research. Although several algorithms for automated classification of presence or absence of dementia have been validated in the Health and Retirement Study (HRS), no such algorithm has yet been applied to the Survey of Health, Ageing and Retirement in Europe (SHARE).

Method: The Langa-Weir classification (LW) was adapted to readily available indicators in SHARE, including immediate and delayed recall. Adapted algorithms additionally included instrumental activities of daily living (IADL) and used cut-offs defined by either sample- or population-level distributions. Performance was compared to logistic and bayesian-logistic regression models and a gradient boosting machine (XGBoost) with the same indicators, adjusting for age groups, gender and educational level. The bayesian-logistic regression used priors for sociodemographic indicators and global dementia incidence. Accuracy, specificity and sensitivity were compared with a train-test split approach in SHARE wave 7 (2017).

Result: In total, N = 72,329 participants (57% female) above age 50 had no missing data on self-reported dementia diagnosis, immediate or delayed recall and IADLs. LW based on immediate and delayed recall with a score cutoff based on dementia population-incidence performed best overall (Accuracy = .92, Balanced Accuracy = .75, Sensitivity = .58, Specificity = .92), and showed greatest similarities to participants with self-reported dementia diagnosis regarding risk factors and comorbidities (i.e., gripstrength, numerical performance, verbal fluency). Results from XGBoost suggested comparable performance however with risk of overfitting.

Conclusion: LW adaptations outperformed regression models regarding sensitivity. Comparisons of risk factor and comorbidity distributions suggest meaningful differences in comorbidities and risk factors in participants classified with and without dementia. With a lack of proxy assessments in SHARE, a suspected healthy volunteer bias and the absence of standardized cognitive assessments, probable dementia detection in SHARE necessarily comes with less confidence compared to algorithms tested in HRS. Nonetheless, performance of LW adaptations in SHARE is in line with previous validation studies in HRS. Future research should validate the algorithms through more extensive cognitive assessments once available.

Table 1. Baseline Characteristics of Participants in the Training and Test Set

	Test Set (N = 68,461)	Training Set (N = 3,868)	P
Age			
Mean (SD)	68.5 (9.62)	69.1 (9.98)	<0.001
Gender			
Female	39,075 (57.1%)	2,214 (57.2%)	.86
Male	29,386 (42.9%)	1,654 (42.8%)	
Educational Level (ISCED 1997)			
Lower Secondary	24,613 (36.0%)	1,485 (38.4%)	.008
Tertiary	18,484 (27.0%)	1,019 (26.3%)	
Upper Secondary	25,364 (37.0%)	1,364 (35.3%)	
Dementia			
Dementia	1,020 (1.5%)	252 (6.5%)	<0.001
No Dementia	67,441 (98.5%)	3,616 (93.5%)	

Note. Reported P-values are based on t-tests for continuous and Chi-squared tests for categorical characteristics.

Table 2. Performance of Classification Algorithms in the Test Set.

	Accuracy	Balanced Accuracy	Sensitivity	Specificity
Algorithm				
LW (statistically informed cutoff)	0.96	0.67	0.36	0.97
LW (incidence-based cutoff)	0.92	0.75	0.58	0.92
LW (including 5 IADLs)	0.96	0.71	0.44	0.97
LW (including 9 IADLs)	0.96	0.72	0.46	0.97
Logistic Regression	0.98	0.63	0.28	0.99
Bayesian Logistic Regression	0.98	0.63	0.27	0.99
XGBoost	0.97	0.65	0.32	0.98

Note. Balanced Accuracy reflects the arithmetic mean of Sensitivity and Specificity. Langa-Weir (LW) adaptations are based on the sum of immediate and delayed recall with and without subtraction according to IADLs (5 IADLs: preparing meals, shopping groceries, making phone calls, taking medication, or managing money; 9 IADLs: 5 IADLs + using a map, doing house/garden work, leaving independently, doing laundry). Unless indicated otherwise, cutoffs for LW adaptations were defined relative to the sample distribution, i.e., statistically informed (z-standardized LW < -2). The incidence-based cutoff relates to the 6.97th LW percentile, based on the pooled prevalence for all-cause dementia in the general population above age 50 [6].

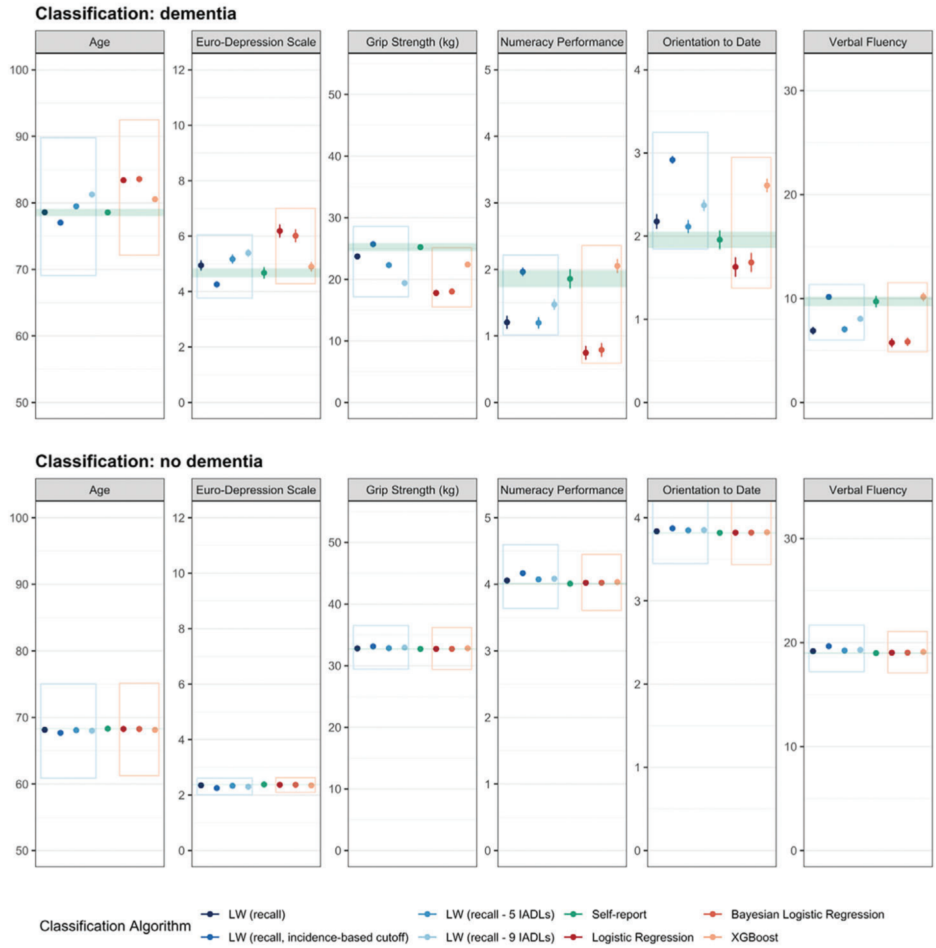
Table 3. Performance of Classification Algorithms in the Training Set.

	Accuracy	Balanced Accuracy	Sensitivity	Specificity
Algorithm				
LW (statistically informed cutoff)	0.93	0.68	0.39	0.97
LW (incidence-based cutoff)	0.89	0.76	0.60	0.91
LW (including 5 IADLs)	0.94	0.74	0.52	0.96
LW (including 9 IADLs)	0.94	0.76	0.55	0.97
Logistic Regression	0.95	0.68	0.37	0.99
Bayesian Logistic Regression	0.95	0.67	0.36	0.99
XGBoost	0.97	0.83	0.66	0.99

Note. Balanced Accuracy reflects the arithmetic mean of Sensitivity and Specificity. Langa-Weir (LW) adaptations are based on the sum of immediate and delayed recall with and without subtraction according to IADLs (5 IADLs: preparing meals, shopping groceries, making phone calls, taking medication, or managing money; 9 IADLs: 5 IADLs + using a map, doing house/garden work, leaving independently, doing laundry). Unless indicated otherwise, cutoffs for LW adaptations were defined relative to the sample distribution, i.e. statistically informed (z-standardized LW < -2). The incidence-based cutoff relates to the 6.97th LW percentile, based on the pooled prevalence for all-cause dementia in the general population above age 50 [6].

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

Figure 1. Comparison of Risk Factors and Comorbidities of Dementia by Algorithmic Probable Dementia Classification (Upper Panel: Classified as Dementia; Lower Panel: Classified as No Dementia) and Self-reported Dementia (Green Line) in the Test Set



Note. Results are reported on original scales, except Grip Strength and Verbal Fluency which were truncated at the 95th percentile. Higher values indicate higher age, higher depressive symptoms and better performance, respectively. Numeracy Performance relates to a subtraction task. Orientation to Date reflects the ability to indicate the day of the month, month, year, and day of the week. Verbal fluency relates to a semantic verbal fluency task with animal naming. Light orange rectangles surround Machine Learning and logistic regression-based algorithms. Light blue rectangles surround Langa-Weir (LW) adaptations based on the sum of immediate and delayed recall with and without subtraction according to IADLs (5 IADLs: preparing meals, shopping groceries, making phone calls, taking medication, or managing money; 9 IADLs: 5 IADLs + using a map, doing house/garden work, leaving independently, doing laundry). Unless indicated otherwise, cutoffs for LW adaptations were defined relative to the sample distribution, i.e., statistically informed (z-standardized LW < -2). The incidence-based cutoff relates to the 6.97th LW percentile, based on the pooled prevalence for all-cause dementia in the general population above age 50 [6]. Error bars indicate the 95% CI. Green shaded areas span over the 95% CI of self-reported dementia classification, suggesting the ground truth.