THE CHALLENGE OF EARLY INTERVENTION: HOW DO YOU IDENTIFY LOW PERFORMANCE?

Erin McMullen Jonaitis, PhD, Rebecca Koscik, PhD, Asenath La Rue, PhD, Bruce P. Hermann, PhD, Mark A. Sager, MD

Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

OBJECTIVE

To develop and compare methods for flagging unexpectedly low or high cognitive performance in those at risk for Alzheimer's disease (AD).

BACKGROUND

Early intervention may slow the development of AD. Little is known, however, about how to identify performance which suggests possible prodromal AD. This study compares methods for identifying longitudinal neurocognitive performance that is different than expected. We hypothesized that cutoffs based on population norms would be less able to detect subtle changes in performance than would regression-based methods (Temkin et al., 1999).

METHODS

554 healthy volunteers (53 ± 6.6 years old at Time 1; 374(67.5%) female; 360(65.0%) college-educated; 548 (98.9%) white; 452(81.6%) FH+) completed a neuropsychological battery at Time 1 and Time 2. For Rey's AVLT delayed recall, we compared three methods for categorizing Time 2 performance:

- For Method 1, we used published norms (Schmidt, 1996; cited in Strauss et al., 2006) to convert raw scores to z-scores.
- For Methods 2 and 3, we used the control group to develop regression models to predict Time 2 performance.
  - Method 2 predicted Time 2 scores using Time 1 scores.
  - Method 3 included Time 1 performance, gender, age, education, season, and test-retest interval in a stepwise fashion (Time 1, interval retained; p<0.1).

For each method, Time 2 scores outside the 90% CI for predicted performance were flagged as worse or better than expected. We repeated these methods for six composite outcomes constructed from our battery (Dowling et al., in press), treating the standardized composite scores from our FH- group as norms for the purpose of Method 1. We examined frequencies of lower-than-expected performance as determined by each of the models.

Dowling M, Hermann BP, La Rue A, Sager MA (In press). *Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease.* Neuropsychology.

Schmidt, M (1996). *Rey Auditory-Verbal Learning Test.* Los Angeles: Western Psychological Services.

Strauss, E, Sherman, EMS, & Spreen, O (2006). *A compendium of neuropsychological tests* (3rd ed.). New York: Oxford University Press.

Temkin, NR, Heaton, RK, Grant, I, & Dikmen, SS (1999). *Detecting significant change in neuropsychological test performance: A comparison of four models.* Journal of the International Neuropsychological Society, 6(3), 362-365.

TABLES & FIGURES

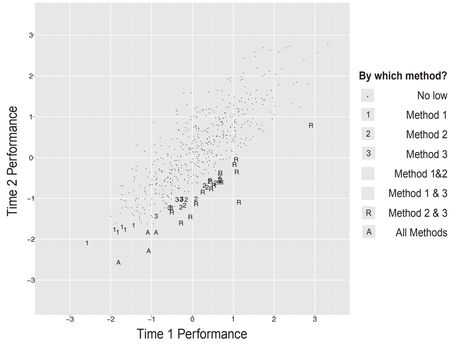
TABLE 1. AVERAGE COGNITIVE SCORES BY METHOD

Means (standard deviation) of seven outcome variables, stratified by classes and method of classification

Tests*	METHOD 1 Norm-Based			METHOD 2 Predictor: Baseline Performance			METHOD 3** Predictors: Baseline & Covariates		
	Worse	Expected	Better	Worse	Expected	Better	Worse	Expected	Better
AVLT –d	2.8 (1.52)	10.4 (2.64)	14.3 (0.86)	7.2 (2.28)	10.9 (2.83)	12.3 (1.96)	6.7 (2.36)	10.7 (2.80)	12.7 (2.30)
Verb	-1.98 (0.18)	0.23 (0.79)	1.84 (0.18)	-0.67 (0.84)	0.24 (0.86)	1.20 (0.60)	-0.75 (0.85)	0.22 (0.86)	0.97 (0.72)
Vis	-2.06 (0.26)	0.28 (0.76)	1.80 (0.14)	-0.71 (0.90)	0.32 (0.84)	0.69 (0.78)	-0.44 (0.82)	0.31 (0.86)	0.65 (0.75)
Spd	-2.03 (0.35)	0.19 (0.78)	2.07 (0.42)	-1.03 (0.86)	0.25 (0.92)	0.66 (1.12)	-0.97 (1.04)	0.20 (0.91)	0.72 (1.03)
Wrkg	-1.93 (0.29)	0.04 (0.80)	2.16 (0.36)	-0.91 (0.68)	0.11 (0.90)	1.48 (0.71)	-1.01 (0.71)	0.09 (0.89)	1.31 (0.74)
Vbl	-2.19 (0.47)	0.19 (0.83)	1.78 (0.06)	-1.45 (0.92)	0.16 (0.96)	1.03 (0.60)	-1.45 (0.92)	0.16 (0.96)	1.03 (0.60)
Imm	-2.09 (0.45)	-0.04 (0.84)	2.13 (0.33)	-1.48 (0.79)	-0.14 (0.88)	1.28 (0.74)	-1.48 (0.79)	-0.14 (0.88)	1.28 (0.74)

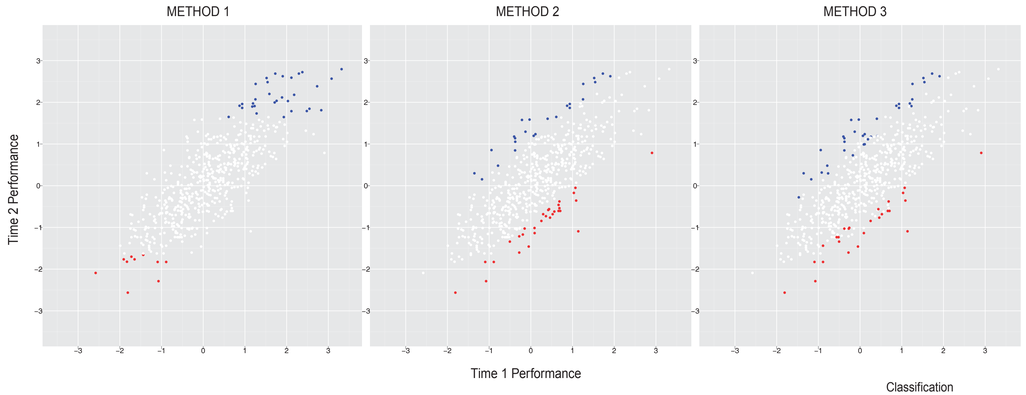
*Test scores range from 0 to 15 for AVLT-d; Composite scores have mean=0 and sd=1.
**Method 3 models included the following predictors in addition to Time 1 performance: AVLT-delay: interval; Verbal ability: season, education; Visuospatial ability: gender, education; Speed & flexibility: age, season; Working memory: season; Verbal learning & Immediate memory, no covariates.

FIGURE 4. METHOD OVERLAP



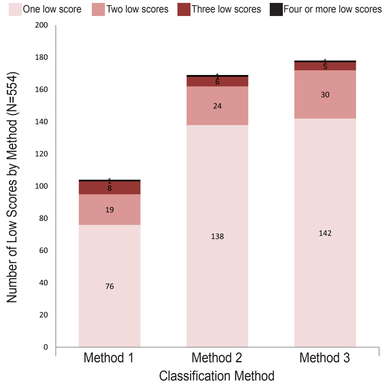
Time 2 vs. time 1 working memory composite scores by method overlap in identifying low performance. No symbol in the legend indicates no overlap in that category.

FIGURES 1-3. CLASSIFYING WORKING MEMORY PERFORMANCE: THREE METHODS



Time 2 vs. time 1 working memory composite scores, by classification method.

FIGURE 5. DOMAIN OVERLAP BY METHOD



The relative proportion of low performers identified by a method that were classified as low on more than one composite score. Numbers represent raw counts of people in each category.

RESULTS

Distribution of classifications: In general, Method 1 identified fewer people as performing worse or better than expected than did either of the other methods. The average percentage (worse, better) across the seven outcomes was (4, 6) for Method 1, (6, 9) for Method 2, and (6, 10) for Method 3. Furthermore, the average performance of those identified by Methods 2 and 3 was less extreme (**Table 1**). This suggests that Methods 2 and 3 allow one to identify more subtle declines.

Model agreement: The classifications made by Models 2 and 3 had good to very good agreement for all outcomes (weighted kappa values .71-.83). However, Model 1 did not agree well with either of the other two (weighted kappa values .08-.38), suggesting that norm-based methods and regression-based methods do not identify the same people as low performers (**Fig. 1-4**).

Overlap of cognitive indicators: Poor performance on one outcome did not generally transfer to poor performance on another outcome. **Figure 5** shows the relative proportion of low performers who were classified as "Worse" on multiple composite score outcomes by Methods 1, 2, or 3.

CONCLUSIONS

As predicted, regression-based methods picked up on variability not captured by norms, which may prove helpful for identifying early decline in those who start out as high performers. However, in our relatively young sample, it is still unclear which of these approaches will do the best job of targeting AD-specific change. Additional follow-up of this sample is needed to determine the overall sensitivity and specificity of these approaches for identifying patients who are on a trajectory to develop MCI and AD. In the meantime, clinicians may want to consider both absolute and relative low performance for a given individual when assessing patient risk of AD.

ACKNOWLEDGEMENTS & CONTACT

This research was supported by the Helen Bader Foundation, Northwestern Mutual Foundation, Extendicare Foundation, NIH grant M01RR03186 (University of Wisconsin Clinical and Translational Research Core) and NIA grant 5R01AG27161-2 (Wisconsin Registry for Alzheimer Prevention: Biomarkers of Preclinical AD).

Erin McMullen Jonaitis, PhD, Wisconsin Alzheimer's Institute
University of Wisconsin School of Medicine and Public Health
Phone: 608-829-3305; Email: jonaitis@wisc.edu; Web: www.wai.wisc.edu