American Psychological Association

Psychological Assessment

ISSN: 1040-3590

2019, Vol. 31, No. 11, 1377–1382
http://dx.doi.org/10.1037/pas0000764

BRIEF REPORT

Scoring Algorithms for a Computer-Based Cognitive Screening Tool: An Illustrative Example of Overfitting Machine Learning Approaches and the Impact on Estimates of Classification Accuracy

Jake Ursenbach, Megan E. O'Connell, and Jennafer Neiser
University of Saskatchewan

Mary C. Tierney
University of Toronto

Debra Morgan, Julie Kosteniuk, and Raymond J. Spiteri
University of Saskatchewan

Computerized cognitive screening tools, such as the self-administered Computerized Assessment of Memory Cognitive Impairment (CAMCI), require little training and ensure standardized administration and could be an ideal test for primary care settings. We conducted a secondary analysis of a data set including 887 older adults (*M* age = 72.7 years, *SD* = 7.1 years; 32.1% male; *M* years education = 13.4, *SD* = 2.7 years) with CAMCI scores and independent diagnoses of mild cognitive impairment (MCI). A study by the CAMCI developers used a portion of this data set with a machine learning decision tree model and suggested that the CAMCI had high classification accuracy for MCI (sensitivity = 0.86, specificity = 0.94). We found similar support for accuracy (sensitivity = 0.94, specificity = 0.94) by overfitting a decision tree model, but we found evidence of lower accuracy in a cross-validation sample (sensitivity = 0.62, specificity = 0.66). A logistic regression model, however, discriminated modestly in both training (sensitivity = 0.72, specificity = 0.80) and cross-validation data sets (sensitivity = 0.69, specificity = 0.74). Evidence for strong accuracy when overfitting a decision tree model and substantially reduced accuracy in cross-validation samples was replicated across 500 bootstrapped samples. In contrast, the evidence for accuracy of the logistic regression model was similar in the training and cross-validation samples. The logistic regression model produced accuracy estimates consistent with other published CAMCI studies, suggesting evidence for classification accuracy of the CAMCI for MCI is likely modest. This case study illustrates the general need for cross-validation and careful evaluation of the generalizability of machine learning models.

Public Significance Statement
Using an archival database, a decision tree machine learning method demonstrated overfitting and had substantially reduced evidence for classification accuracy measured in cross-validation, but the statistical method results in similar evidence of classification in training and cross-validation samples. The evidence for classification accuracy of the Computerized Assessment of Mild Cognitive Impairment for cognitive impairment is modest, and this has clinical implications.

Keywords: mild cognitive impairment, accuracy, computerized testing, machine learning

This article was published Online First August 15, 2019.

Jake Ursenbach and Megan E. O'Connell, Department of Psychology, University of Saskatchewan; Jennafer Neiser, Department of Computer Science, University of Saskatchewan; Mary C. Tierney, Department of Family & Community Medicine, University of Toronto; Debra Morgan and Julie Kosteniuk, Canadian Centre for Health and Safety in Agriculture, University of Saskatchewan; Raymond J. Spiteri, Department of Computer Science, University of Saskatchewan.

Jake Ursenbach received a Master's Award from the Canadian Institutes for Health Research. Further trainee funding was provided by the Canadian Consortium on Neurodegeneration in Aging (CCNA). CCNA is supported by a grant from the Canadian Institutes of Health Research with funding from several partners, including the Saskatchewan Health Research Foundation. Mary Catherine Tierney was supported by a Clinician Scientist Award from the Department of Family and Community Medicine, University of Toronto and Sunnybrook Health Sciences Centre.

Correspondence concerning this article should be addressed to Megan E. O'Connell, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada. E-mail: megan.oconnell@usask.ca

Pathophysiological changes due to Alzheimer's disease (AD) accumulate in the brain for years prior to symptom onset and typically produce a gradual decline in cognitive functioning. Evidence of a decline in cognitive functioning is a core feature of mild cognitive impairment (MCI; Petersen et al., 2018) and dementia (McKhann et al., 2011). Obtaining information about a patient's objective cognitive functioning is often difficult in primary care, and it is important that cognitive screening tests are administered correctly (i.e., in a standardized manner) to ensure patient performance can be appropriately compared to the published cutoff scores. Many primary care providers have reported barriers to making a dementia diagnosis, including a lack of assessment instruments, limited knowledge of assessment tool administration, and limited time to carry out cognitive testing during appointments (Aminzadeh, Molnar, Dalziel, & Ayotte, 2012; Bradford, Kunik, Schulz, Williams, & Singh, 2009; Iracleous et al., 2010). Computerized cognitive screening instruments are a promising way to improve timely dementia diagnosis in primary care because instruments are often self-administered: Patients can complete them relatively independently, and no staff training for standardized administration is necessary. Computerized cognitive tests are also typically self-scored, thus reducing errors and saving time (Zygouris & Tsolaki, 2015).

A range of computerized cognitive tests has been used to assess cognitive deficits in older adults (Zygouris & Tsolaki, 2015). In a recent systematic review, the Computer Assessment of Mild Cognitive Impairment (CAMCI) was reported to have high sensitivity and specificity for MCI, with the caveat of a lack of replication across samples (Aslam et al., 2018). Saxton et al. (2009) reported evidence that the CAMCI discriminated well between participants with MCI and those without (sensitivity = 0.86, specificity = 0.94), which is higher than the estimates obtained by others for the CAMCI (Becker et al., 2011; Nieto, Albert, Morrow, & Saxton, 2008; Tierney et al., 2014). The CAMCI has been used in various clinical populations (e.g., vascular risk, Williams, Park, Tsang, Sperling, & Manning, 2018; HIV, Becker et al., 2011; Rosenthal et al., 2013; Valcour, Paul, Chiao, Wendelken, & Miller, 2011; cardiac arrest, Koller et al., 2017; Sabedra et al., 2015; knee osteoarthritis, Morone, Abebe, Morrow, & Weiner, 2014). Finally, the CAMCI demonstrated a positive impact on primary care providers' dementia-related care decisions (Tierney, Charles, Naglie, Jaakkimainen, & Moineddin, 2017; Tierney et al., 2014).

Saxton et al. (2009) used a classification and regression tree model, which is a data-mining approach that involves creating a tree-like structure in which a set of predictors (one of the 32 possible CAMCI test scores) is tested at each level; the predictor that best classifies cases is used as a decision point on that level of the tree (Breiman, Friedman, Olshen, & Stone, 1984). Saxton et al. (2009) reported that they used SPSS and "derived the sensitivity and specificity from the final tree model which was created using 10-fold cross-validation" (p. 5). SPSS only reports the decision tree built on the final model and reports the classification accuracy of this final model (IBM Support, 2016). We argue that the SPSS cross-validation risk obtained from the 10-fold procedure was not reported by Saxton et al. (2009); rather, the risk from the model built and evaluated on the full sample was reported. We postulated the decision tree model described might have been evaluated on the same data set with which it was trained, resulting in overfitting. Overfitting occurs when a model relies on error in the training data to increase predictive accuracy, resulting in a "gap between the training error and the test error [that] is too large" (Goodfellow, Bengio, & Courville, 2016, p. 109). We explored the accuracy statistics in a similar data set to the one used by Saxton et al. (2009) using a similar decision tree method to allow us to determine whether the accuracy reported by Saxton et al. (2009) was inflated due to overfitting. We also developed an alternative algorithm that may be used with the CAMCI scores to estimate a patient's probability of cognitive impairment.

Method

Participants

This study is a secondary analysis of a data set of adults living in western Pennsylvania from Psychological Software Tools. Inclusion criteria were described in Saxton et al. (2009), but briefly, the inclusion criteria included English-speaking adults with no evidence of neurological conditions. In addition to the CAMCI, participants completed a pencil-and-paper neuropsychological assessment. The archival data set consisted of a total of 1,173 participants. In total, 286 participants were excluded because they did not receive an adjudicated classification of MCI or no cognitive impairment from two clinicians based on the paper-and-pencil neuropsychological data. Consistent with the procedure described by Saxton et al. (2009), we removed participants with cognitive performance consistent with a conceptualization of dementia defined as greater than 2 standard deviations below age-adjusted norms on two domains, one of which was memory ($N = 13$), for a final sample size of 887. Participants in the data set were administered either the original version ($n = 589$) or the alternate form ($n = 298$). The local ethical review panel approved the secondary analysis of this previously collected data.

Data Collection and Measures

Details of the CAMCI are provided by Saxton et al. (2009), but briefly the CAMCI assesses five cognitive domains using eight tasks: attention assessed with visual distractor and forward digit span tasks, working memory with a backward digit span task, verbal memory assessed by free recall and a recognition task, visual memory with a recognition task, and executive function assessed with a go/no go task. In addition, a virtual environment task simulates a shopping trip while measuring recognition memory, incidental memory, and prospective memory.

Statistical Analysis

To replicate the decision tree analysis used by Saxton et al. (2009), we used the decision tree approach in the *rpart* package (Therneau & Atkinson, 2018) in R (Version 3.4.2; R Core Team, 2017) with "a gini impurity measure, a tree depth of 13, 10 minimum cases in parent node, 1 minimum case in child node, and minimum change in improvement at a Gini impurity level of 0.0001" (Saxton et al., 2009, p. 5). The logistic regression model was built using raw data with the "glm" function in R optimized using backward and forward stepwise model building to maximize the Akaike information criterion.

The decision tree and logistic regression models were developed on 524 cases, referred to as the *training subsample,* and tested in the remaining 363 cases, referred to as the *cross-validation subsample*. The Saxton et al. (2009) model was created on a sample of 524 cases; consequently, we chose this same sample size for the training subsample with all remaining cases used for cross-validation. We used bootstrapping (samples each of $N = 887$ created with $n = 524$ training to develop the decision tree and logistic regression models that were tested in cross-validation subsamples of $n = 363$; repeated for 500 trials) to allow us to explore whether results were dependent on the specific cases chosen for the *training* versus *cross-validation* subsamples.

We estimated the probability of impairment of each case using each model in order to evaluate the range of possible cutoff points for predictive accuracy. In the receiver operating characteristic (ROC) curves, we display the average sensitivity and specificity at each cutoff point across the 500 bootstrapped samples, and we obtained the 95% confidence interval of the sensitivity and specificity for each cutoff point obtained from the empirical distribution of the 500 bootstrapped samples. We calculated point estimates for the optimal sensitivity and specificity for each model evaluating a range of cutoffs using Youden's (1950) criterion, with *maximum* (*sensitivity* + *specificity*). We reported the average sensitivity and specificity of the bootstrapped samples and the 95% confidence intervals (obtained from the bootstrapped empirical distribution).

Results

The data set consisted of demographic information and 32 variables recorded during test administration. Two variables were removed because observations were missing for over 85% of participants (median reaction time [RT] for false positives on a simple RT task and median RT for correct responses on first repeat trials on a recurring-pictures task), while the remaining 30 variables included observations for at least 90% of participants. Participant characteristics are presented in Table 1.

As can be seen in Figure 1, the decision tree model to replicate the results of Saxton et al. (2009) using a similar data set demonstrated a high degree of predictive accuracy on the *training sample* (point estimate of sensitivity 0.94, 95% CI [0.87, 0.97]; point estimate of specificity 0.94, 95% CI [0.89, 0.97]) but subsequently dropped when predicting impairment in the *cross-validation sample* (point estimate of sensitivity 0.62, 95% CI [0.50, 0.74]; point estimate of specificity 0.66, 95% CI [0.52, 0.75]). Moreover, the confidence interval obtained from the empirical distribution of the sensitivity and specificity of the 500 bootstrapped samples was wide for the decision tree model in the cross-validation sample. In contrast, the logistic regression model provided more modest evidence for classification accuracy measured with sensitivity and specificity, but the classification accuracy was similar in the training data set (point estimate of sensitivity 0.72, 95% CI [0.68, 0.76]; point estimate of specificity 0.80, 95% CI [0.77, 0.82]) and the cross-validation samples (i.e., the confidence intervals overlapped; point estimate of sensitivity 0.69, 95% CI [0.63, 0.74]; point estimate of specificity 0.74, 95% CI [0.71, 0.78]).

Table 1
Demographic and Clinical Characteristics of Participants in the Computerized Assessment of Memory Cognitive Impairment Data Set Reported as Mean (Minimum–Maximum; SD) or Frequency Counts

Variable	Sample
N	887
Sex, male/female, *n*	291/596
Age (range; *SD*)	72.7 (60–93; 7.1)
Education (range; *SD*)	13.4 (7–20; 2.7)
Adjudicated classification, normal/impaired, *n*	559/328
Mini-mental state examination (range; *SD*)	28.1 (21–30; 1.7)

Discussion

We argued that it was possible the classification algorithm reported by Saxton et al. (2009) was overfit to the data. In support of this, we were able to replicate their sensitivity and specificity only when we trained and evaluated the decision tree model on the same data set, resulting in overfitting. We used bootstrapping and found the high accuracy in the training data using a "decision tree" method for a classification algorithm substantially dropped in accuracy when cross-validated irrespective of the specific data used for *training* versus for *cross-validation*. In contrast, the logistic regression model from the 32 CAMCI raw scores was able to predict cognitive impairment in the *training* and *cross-validation* samples with a similar degree of accuracy. The evidence for the accuracy of the CAMCI presented in this article in the cross-validation samples was lower than that reported by Saxton et al. (2009; sensitivity of 0.86, specificity of 0.94) and Becker et al. (2011; sensitivity of 0.72, specificity of 0.97), but the logistic regression model's evidence for accuracy is comparable to that reported by Tierney et al. (2014; sensitivity of 0.80, specificity of 0.74) and Nieto et al. (2008; sensitivity of 0.79, specificity of 0.67). These differences can likely be attributed to sampling error and variation in the reference standard used to classify MCI versus no cognitive impairment. For example, the reference standard used by Nieto et al. (2008) was unclear, Becker et al. (2011) used a composite score from a battery of pencil-and-paper neuropsychological tests, and Tierney et al. (2014) defined cognitive impairment on the neuropsychological tests as two or more test scores more than 1.5 standard deviations from the normative mean of the test (the same definition used by Saxton et al., 2009).

A comparison of the evidence for accuracy from the decision tree approach in the cross-validation samples reveals that the estimate of sensitivity or specificity includes 50% (0.50) or chance classification in the confidence interval. In contrast, the estimate of sensitivity and specificity for the logistic regression approach to summarizing cognition across the battery of tests does not include .50, suggesting more robust support for accuracy of any interpretation of overall cognitive impairment from the CAMCI using the logistic regression algorithm.

These results are an illustrative example of the potential pitfalls of using packaged software for machine learning approaches that are not commonly used in psychological research and underscore the well-established need in psychometrics for cross-validation. Decision tree models are ideally suited for automatically summarizing a large amount of information to a single categorical variable without making parametric assumptions, such as from a neuropsychological battery for classifying impairment, which is

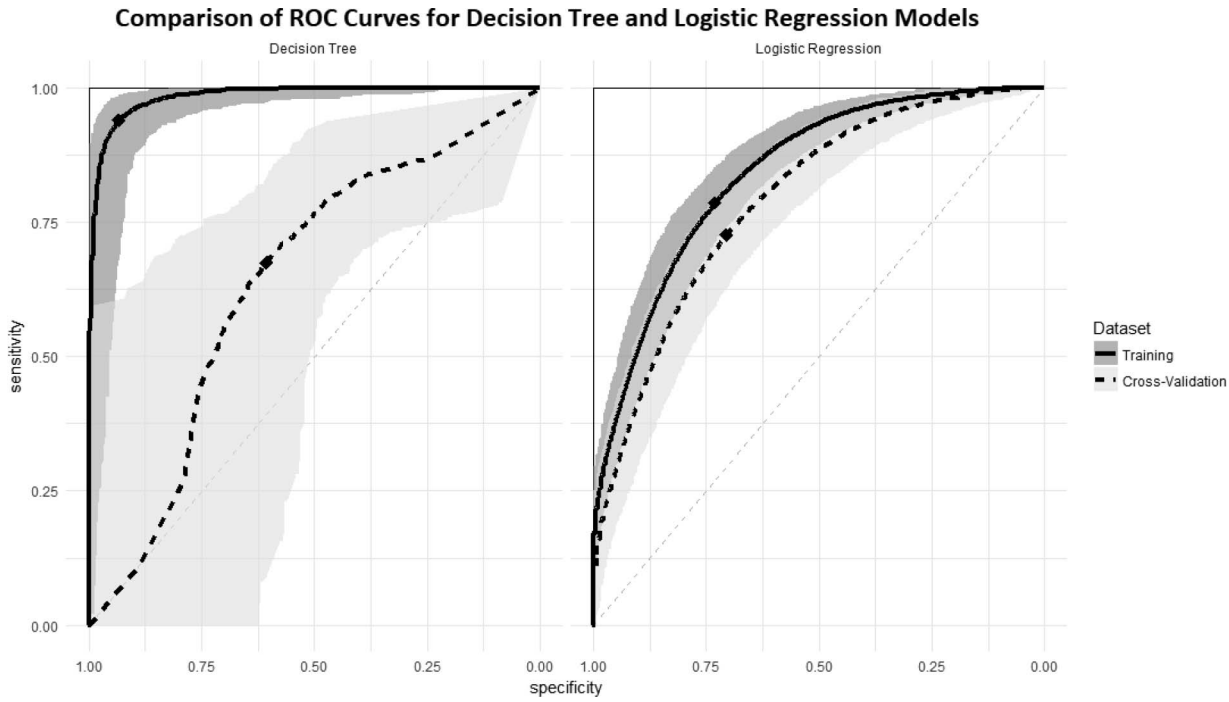

Figure 1. Across the 500 bootstrapped samples of testing and cross-validation subsamples, the median receiver operating characteristic (ROC) curve is indicated with bold lines, and the 95% confidence interval obtained from the empirical distribution is indicated by the shaded area for the training (dark) and cross-validation (light) subsamples with the decision tree and logistic regression models. Points indicate optimal sensitivity and specificity according to Youden's formula. The light-gray dotted line indicates a 50% chance to classify a case correctly.

typically the role of the clinical neuropsychologist during clinical interpretation. It is notable that the logistic regression approach provided similar results in the training and cross-validation samples, supporting the use of logistic regression to summarize scores from a neuropsychological battery to aid in classification. One drawback, however, of logistic regression versus the decision tree approach is the limitations of logistic regression for large numbers of variables. Preprocessing of data (e.g., with the use of principal components to reduce the number of predictors to enter into a logistic regression; Harrell, 2015) would mitigate this concern and is likely best practice. Given the large number of potential predictors (recall the CAMCI included 32 possible scores), it is difficult to justify a priori which predictors should be modeled without a strong theoretical basis. The risk of overfitting increases significantly when one models predictors based on "data mining" approaches (e.g., screening based on univariate associations) unless one accounts for the additional "hidden" degrees of freedom (Harrell, 2015). In future research, a short-form version of the CAMCI could be developed using "model approximation" (Harrell, 2015), specifically by using the predicted risk scores from the logistic regression model as the dependent variable and CAMCI raw scores as predictors. Then a backward stepwise regression method could be used with R^2 values calculated for each iteration. An arbitrary cutoff of 0.95 could be used, and predictors could be dropped until the second model accounted for only 95% of the variance of the first model. The resulting model would be more parsimonious and likely require only a subset of the tasks.

Overfitting can be identified based on a large gap between training and testing error (Goodfellow et al., 2016). When it is identified, it can be addressed using regularization. Regularization is "any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error" (Goodfellow et al., 2016, p. 224). For example, lasso regression uses a penalized estimation method that constrains regression coefficients so their sum is below a constant that is chosen through cross-validation. As a result, variables that are poorer predictors of the criterion have their coefficients reduced toward zero, thereby reducing variance in the model that may be spuriously related to the criterion (Harrell, 2015).

These analyses are limited by the use of raw CAMCI scores versus normatively corrected scores, but this applies to both approaches and hence does not detract from the comparison of processing the same data with different methods. Arguably, when transforming results from a neuropsychological battery into a summary classification of cognitive impairment versus no impairment, use of demographic corrections for variables such as age could be cautioned. In a simulation study subsequently tested in a clinical sample, O'Connell and Tuokko (2010) demonstrated that age corrections reduced sensitivity when the condition of interest for classification (e.g., MCI or dementia) is also associated with age. They argue that across a neuropsychological battery where test scores are interpreted independently, such as would be performed by a clinical neuropsychologist, the reduced sensitivity is balanced by the battery approach (which increases sensitivity and

reduces specificity, thus the evidence-based practice of adjusting for spurious impairments suggested by Crawford, Garthwaite, & Gault, 2007). When a single score is used, however, corrections for demographic variables such as age, education, and sex would be cautioned (O'Connell & Tuokko, 2010). This could be similarly argued for decision tree or logistic regression approaches that summarize neuropsychological batteries into a single classification score to identify cognitive impairment versus no cognitive impairment. Finally, our data are limited by the unequal representation based on sex, with a larger proportion of females in the sample. Nonbootstrapped analyses with logistic regression models performed separately based on sex suggested the models resulted in similar estimates of sensitivity and specificity for males and females, which mitigates some of the concern about the unequal representation of females in the sample used to create the logistic regression model. Nevertheless, replication of the logistic regression model in varied clinical samples is needed.

In summary, we provided evidence that the accuracy of the scoring algorithm reported by Saxton et al. (2009) was likely an overestimate, and therefore, health care providers should reference the cumulative evidence for the diagnostic accuracy studies of the CAMCI. These data are also informative in their comparison of two different methods to develop algorithms to summarize cognitive performance from a battery of cognitive tests. A statistical method based on logistic regression to create a predictive model resulted in similar evidence of classification for normal cognition versus MCI in the training and the cross-validation samples. In contrast, the decision tree method demonstrated substantially reduced evidence for classification accuracy in the cross-validation samples versus the training sample. In addition, the confidence intervals for the decision tree method in the training data were large, and, importantly, the wide confidence interval included chance classification. These data are also an important reminder that cross-validation is particularly critical to examine when evaluating machine learning models, such as decision tree models.

References

Aminzadeh, F., Molnar, F. J., Dalziel, W. B., & Ayotte, D. (2012). A review of barriers and enablers to diagnosis and management of persons with dementia in primary care. *Canadian Geriatrics Journal, 15,* 85–94. http://dx.doi.org/10.5770/cgj.15.42

Aslam, R. W., Bates, V., Dundar, Y., Hounsome, J., Richardson, M., Krishan, A., . . . Sikdar, S. (2018). A systematic review of the diagnostic accuracy of automated tests for cognitive impairment. *International Journal of Geriatric Psychiatry, 33,* 561–575. http://dx.doi.org/10.1002/gps.4852

Becker, J. T., Dew, M. A., Aizenstein, H. J., Lopez, O. L., Morrow, L., & Saxton, J. (2011). Concurrent validity of a computer-based cognitive screening tool for use in adults with HIV disease. *AIDS Patient Care and STDs, 25,* 351–357. http://dx.doi.org/10.1089/apc.2011.0051

Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P., & Singh, H. (2009). Missed and delayed diagnosis of dementia in primary care: Prevalence and contributing factors. *Alzheimer Disease and Associated Disorders, 23,* 306–314. http://dx.doi.org/10.1097/WAD.0b013e3181a6bebc

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York, NY: Chapman & Hall/CRC.

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21,* 419–430. http://dx.doi.org/10.1037/0894-4105.21.4.419

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. Retrieved from http://www.deeplearningbook.org

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis* (2nd ed.). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-19425-7

IBM Support. (2016). *What type of cross-validation is used in SPSS classification trees?* Retrieved June 19, 2018, from http://www.ibm.com/support,//www.ibm.com/support/docview.wss?uid=swg21478363

Iracleous, P., Nie, J. X., Tracy, C. S., Moineddin, R., Ismail, Z., Shulman, K. I., & Upshur, R. E. (2010). Primary care physicians' attitudes towards cognitive screening: Findings from a national postal survey. *International Journal of Geriatric Psychiatry, 25,* 23–29.

Koller, A. C., Rittenberger, J. C., Repine, M. J., Morgan, P. W., Kristan, J., & Callaway, C. W., & the Post-Cardiac Arrest Service. (2017). Comparison of three cognitive exams in cardiac arrest survivors. *Resuscitation, 116,* 98–104. http://dx.doi.org/10.1016/j.resuscitation.2017.04.011

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7,* 263–269. http://dx.doi.org/10.1016/j.jalz.2011.03.005

Morone, N. E., Abebe, K. Z., Morrow, L. A., & Weiner, D. K. (2014). Pain and decreased cognitive function negatively impact physical functioning in older adults with knee osteoarthritis. *Pain Medicine, 15,* 1481–1487. http://dx.doi.org/10.1111/pme.12483

Nieto, M. L., Albert, S. M., Morrow, L. A., & Saxton, J. (2008). Cognitive status and physical function in older African Americans. *Journal of the American Geriatrics Society, 56,* 2014–2019. http://dx.doi.org/10.1111/j.1532-5415.2008.01938.x

O'Connell, M. E., & Tuokko, H. (2010). Age corrections and dementia classification accuracy. *Archives of Clinical Neuropsychology, 25,* 126–138. http://dx.doi.org/10.1093/arclin/acp111

Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., . . . Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. *Neurology, 90,* 126–135. http://dx.doi.org/10.1212/WNL.0000000000004826

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosenthal, L. S., Skolasky, R. L., Moxley, R. T., IV, Roosa, H. V., Selnes, O. A., Eschman, A., . . . Sacktor, N. (2013). A novel computerized functional assessment for human immunodeficiency virus-associated neurocognitive disorder. *Journal of Neurovirology, 19,* 432–441. http://dx.doi.org/10.1007/s13365-013-0195-5

Sabedra, A. R., Kristan, J., Raina, K., Holm, M. B., Callaway, C. W., Guyette, F. X., . . . the Post Cardiac Arrest Service. (2015). Neurocognitive outcomes following successful resuscitation from cardiac arrest. *Resuscitation, 90,* 67–72. http://dx.doi.org/10.1016/j.resuscitation.2015.02.023

Saxton, J., Morrow, L., Eschman, A., Archer, G., Luther, J., & Zuccolotto, A. (2009). Computer assessment of mild cognitive impairment. *Postgraduate Medicine, 121,* 177–185. http://dx.doi.org/10.3810/pgm.2009.03.1990

Therneau, T. M., & Atkinson, E. J. (2018). *An introduction to recursive partitioning using the RPART routines*. Retrieved from https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Tierney, M. C., Charles, J., Naglie, G., Jaakkimainen, L., & Moineddin, R. (2017). The effects of computerized cognitive testing of older patients on primary care physicians' approaches to care: A Canadian study. *Alzhei-*

mer Disease and Associated Disorders, 31, 62–68. http://dx.doi.org/10.1097/WAD.0000000000000159

Tierney, M. C., Naglie, G., Upshur, R., Moineddin, R., Charles, J., & Jaakkimainen, R. L. (2014). Feasibility and validity of the self-administered computerized assessment of mild cognitive impairment with older primary care patients. *Alzheimer Disease and Associated Disorders, 28,* 311–319. http://dx.doi.org/10.1097/WAD.0000000000000036

Valcour, V., Paul, R., Chiao, S., Wendelken, L. A., & Miller, B. (2011). Screening for cognitive impairment in human immunodeficiency virus. *Clinical Infectious Diseases, 53,* 836–842. http://dx.doi.org/10.1093/cid/cir524

Williams, I. C., Park, M. H., Tsang, S., Sperling, S. A., & Manning, C. (2018). Cognitive function and vascular risk factors among older African American adults. *Journal of Immigrant and Minority Health, 20,* 612–618. http://dx.doi.org/10.1007/s10903-017-0583-7

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35. http://dx.doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zygouris, S., & Tsolaki, M. (2015). Computerized cognitive testing for older adults: A review. *American Journal of Alzheimer's Disease & Other Dementias, 30,* 13–28. http://dx.doi.org/10.1177/1533317514522852

Received November 12, 2018
Revision received June 17, 2019
Accepted July 8, 2019 ■