J Head Trauma Rehabil
Vol. 34, No. 3, pp. E64–E74

Using Decision Tree Methodology to Predict Employment After Moderate to Severe Traumatic Brain Injury

Katharine A. Stromberg, BS; Amma A. Agyemang, PhD; Kristin M. Graham, PhD; William C. Walker, MD; Adam P. Sima, PhD; Jennifer H. Marwitz, MA; Cynthia Harrison-Felix, PhD; Jeanne M. Hoffman, PhD; Allen W. Brown, MD; Jeffrey S. Kreutzer, PhD; Randall Merchant, PhD

Objective: To build decision tree prediction models for long-term employment outcomes of individuals after moderate to severe closed traumatic brain injury (TBI) and assess model accuracy in an independent sample. **Setting:** TBI Model Systems Centers. **Participants:** TBI Model Systems National Database participants injured between January 1997 and January 2017 with moderate to severe closed TBI. Sample sizes were 7867 (year 1 postinjury), 6783 (year 2 postinjury), and 4927 (year 5 postinjury). **Design:** Cross-sectional analyses using flexible classification tree methodology and validation using an independent subset of TBI Model Systems National Database participants. **Main Measures:** Competitive employment at 1, 2, and 5 years postinjury. **Results:** In the final employment prediction models, posttraumatic amnesia duration was the most important predictor of employment in each outcome year. Additional variables consistently contributing were age, preinjury education, productivity, and occupational category. Generally, individuals spending fewer days in posttraumatic amnesia, who were competitively employed preinjury, and more highly educated had better outcomes. Predictability in test data sets ranged from a C-statistic of 0.72 (year 5; confidence interval: 0.68-0.76) to 0.77 (year 1; confidence interval: 0.74-0.80). **Conclusion:** An easy-to-use decision tree tool was created to provide prognostic information on long-term competitive employment outcomes in individuals with moderate to severe closed TBI. Length of posttraumatic amnesia, a clinical marker of injury severity, and preinjury education and employment status were the most important predictors. **Key words:** *postinjury employment*, *posttraumatic amnesia*, *prognostic model*, *traumatic brain injury*

TRAUMATIC BRAIN INJURY (TBI) is a leading cause of disability among adults in the United States, with estimates of more than 1.1% of the population living with long-term or lifelong TBI-related disability.[1] Because of the persisting and pervasive impairments associated with TBI and young average injury age, postinjury employment remains a critical touchstone. The benefits of employment go beyond providing financial support; competitive employment also promotes psychosocial health, including positive self-esteem and facilitating broader community reintegration.[2–4]

Current evidence demonstrates that preinjury employment is a leading determinant of postinjury employment after moderate to severe TBI.[5,6] Using the TBI Model Systems (TBIMS) National Database (NDB),

Author Affiliations: *Departments of Biostatistics (Ms Stromberg and Dr Sima) and Physical Medicine and Rehabilitation (Drs Agyemang, Graham, Walker, Kreutzer, and Merchant and Ms Marwitz), Virginia Commonwealth University, Richmond; The Traumatic Brain Injury Model Systems National Data and Statistical Center, Craig Hospital, Englewood, Colorado (Dr Harrison-Felix); Department of Rehabilitation Medicine, University of Washington, Seattle (Dr Hoffman); and Department of Physical Medicine and Rehabilitation, Mayo Clinic, Rochester, Minnesota (Dr Brown).*

This study was supported by National Institute of Health (NIH) no. R21HD089097. The data contents for this manuscript were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90 DP0013). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). Postdoctoral fellows were supported by NIDILRR grant number 90AR5025.

The contents of this manuscript do not necessarily represent the policy of NIH or NIDILRR, ACL, and HHS, and the endorsement by the Federal Government should not be assumed.

The authors declare no conflicts of interest.

Corresponding Author: *Adam P. Sima, PhD, Department of Biostatistics, Virginia Commonwealth University, Box 980032, Richmond, VA 23298 (adam.sima@vcuhealth.org).*

DOI: 10.1097/HTR.0000000000000438

Walker et al[6] found that survivors who were employed preinjury were nearly 3 times more likely to be employed 1 year postinjury. In addition, individuals with professional or managerial jobs preinjury had the highest odds of returning to work post-TBI compared with those with skilled or manual labor jobs. Similarly, Cuthbert et al[5] found that preinjury vocational status was associated with unemployment and part-time employment 2 years post-TBI, such that individuals who were students and unemployed preinjury were 56% and 40% more likely to be unemployed postinjury than those employed preinjury. This is consistent with other acquired disabilities such as spinal cord injury, where postinjury employment is also heavily predicated on preinjury occupation.[7]

Other variables shown to be predictive of employment after TBI are education, age, and inpatient rehabilitation length of stay (LOS).[8] Keyser-Marcus and colleagues found that, besides preinjury productivity, younger age and higher education were associated with increased rates of postinjury employment at 1 year follow-up, and age remained a significant predictor at 2- and 3-year follow-up.[8] Literature has also shown that driving status and marital status are important predictors of employment stability.[9–11] In addition, prior alcohol and drug use has been shown to be associated with lower rates of return to productive employment after TBI.[12]

The existing literature in this area has limitations. First, many studies including individuals with moderate and severe TBI performed outside of the TBIMS tend to have small sample sizes.[2,13,14] Most also lack evidence of reproducibility, having not been validated in independent data sets, raising doubts about generalizability. Past studies have also mainly focused on outcomes within 1 year of injury,[13–15] leaving unanswered important questions about the longer-term employment prospects for survivors. These limitations lead to models that translate poorly in real-world clinical applications.

Classification tree methodology is ideal for building clinically useful predictive models because it uses simple logic for classifying patients, making it easy for patients and clinical specialists to use.[16] Decision tree analysis is similar to the clinical pathways that physicians are accustomed to working with, making it an attractive modeling choice for disseminating practical models.[17,18] Classification trees are created by starting with a large data set, and the variables that are most highly related to the outcome are chosen to be included.[19] Besides increased simplicity and practicality of models, classification tree analysis allows for the estimation of predicted outcomes, unlike the parameter estimates or odds ratios produced by traditional logistic regression models.[20] These methods also automatically include interactions as opposed to them having to be entered manually into a regression model and they overcome any collinearity between variables.[19] These methods are also preferred to other ensemble methods, such as random forests, not only because there is limited clinical utility in those models but also because they cannot take into account missing data, which could limit sample sizes dramatically.[18]

Surprisingly, classification trees are rarely used for TBI prognostic modeling or employment outcomes. One of the few publications that used classification trees to model postinjury employment used the Rehabilitation Service Administration (RSA-911) data set.[21] The sample was 66% male and 78% white; most notably, the study invoked the χ^2 automatic interaction detector, which has several limitations.[21] The χ^2 automatic interaction detector does not involve any type of automated pruning, leading to overfitted models. It also cannot handle missing values which can limit sample sizes.[22] An earlier publication, employing classification trees using TBIMS NDB data, considered only binary splits of variables with only 1-year outcomes analyzed.[13] Allowing multiway splitting maximizes the use of the data and improves predictability.[16]

For outcomes as crucial as employment, families and individuals with TBI want and need practical prognostic information spanning multiple horizons. To create a clinically useful tool, decision tree theory employing multiway, data-based splitting was used to create models to predict employment at 1, 2, and 5 years post-TBI. Specifically, the goals of this study were to use the TBIMS NDB, a multicenter database with thousands of participants, to create a user-friendly prognostic model for long-term employment outcomes, based on patient characteristics available upon rehabilitation discharge, in individuals who incurred a moderate to severe closed TBI and to assess the predictability of these models using independent data sets.

METHODS

Participants

All participants were consented and enrolled in the TBIMS NDB, funded by the National Institute on Disability, Independent Living, and Rehabilitation Research.[23] All centers were granted institutional review board approval. Eligible TBIMS participants were (1) 16 years of age or older at injury, (2) admitted to a TBIMS acute care hospital within 72 hours of injury, (3) received care through a TBIMS center, and (4) sustained TBI with any of the following: Glasgow Coma Scale score at emergency admission of less than 13 (not due to intoxication, intubation, or sedation); loss of consciousness, not due to sedation or intoxication, for more than 30 minutes; posttraumatic amnesia (PTA) for more than 24 hours; or injury-related computed tomographic findings.

Given the focus on employment, participants younger than 18 years or older than 60 years at injury, who were more likely to be unemployed because they were students or retired, were excluded from these analyses, as were those with outcomes of death or vegetative state. In addition, due to differing variable definitions prior to 1997,[24] participants with injury dates only between January 1997 and 2017 were included. Other exclusion criteria were penetrating TBI, due to known outcome differences, missing employment outcome, and missing PTA duration data. Since we conducted several cross-sectional analyses, participants may have been included in one analysis but not another (see Figure 1). "Excluded" participants were excluded from all further analyses because of death, insufficient time passing since injury, or a withdrawn consent. Conversely, "removed" participants may have been included in subsequent analyses but for the given cross-sectional analysis had missing employment status data, were in a vegetative state postinjury, refused to participate, were incarcerated, or were participating in a site that lost funding. The sample sizes in Figure 1 are calculated by starting with the total (14 479) and subtracting the "not eligible" and the "excluded" and "removed" subjects at each year. The final sample sizes used in analyses were 7867, 6783, and 4927 for follow-up periods 1, 2, and 5 years postinjury, respectively.

Outcomes

The primary outcome was current competitive employment, defined as a full or part-time job paying at least minimum wage and not in supported employment.[25]

Predictors

Model predictors were selected on the basis of previous research showing their relevance to postinjury employment. Demographic variables considered were baseline sex, age at injury, and education level (less than high school, high school/GED, more than high school). Race was not included because the variety of racial groupings could complicate classifications and limit clinical usefulness of the prognostic models.

Preinjury health characteristics included history of prior TBIs both number and categorical (yes/no), based on The Ohio State University TBI Identification interview,[26,27] history of problem alcohol use (yes/no),[28] and history of illicit drug use (yes/no). For participants injured from October 1999 to present, alcohol abuse was based on the Behavioral Risk Factors Surveillance System questions,[29] while for those injured before October 1999, alcohol abuse was determined by

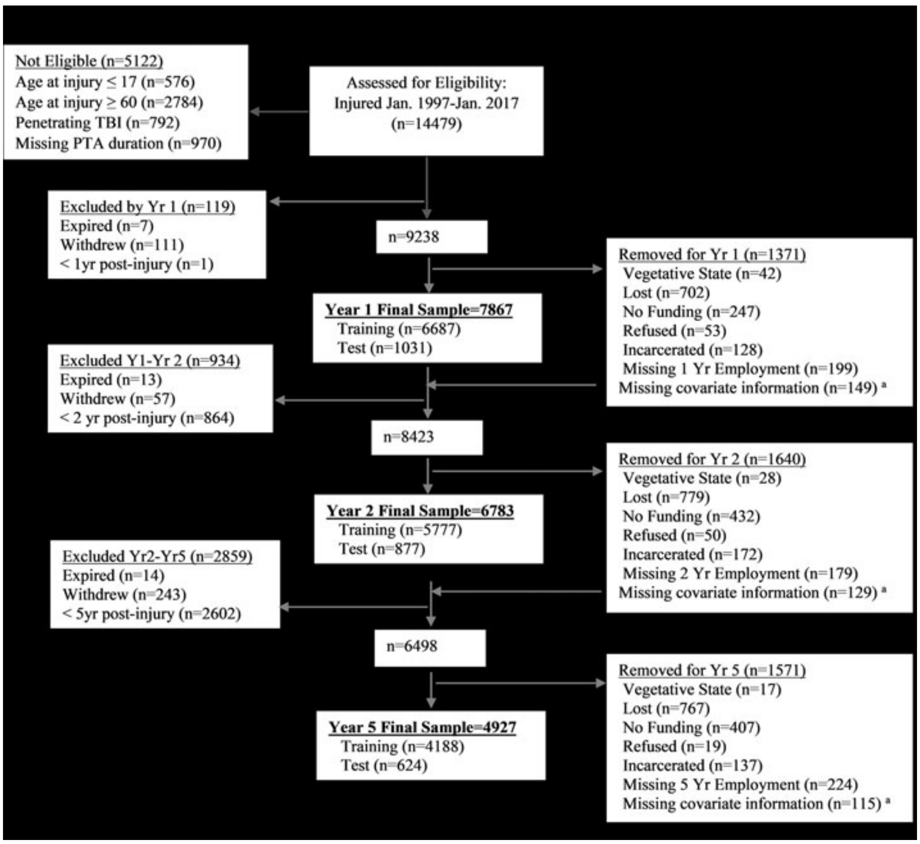

Figure 1. CONSORT diagram.

the Quantity-Frequency-Variability Index.[30] Preinjury socioeconomic characteristics included occupational category (Professional, Skilled, Manual, Unemployed) and productivity (yes/no). Being productive included being competitively employment, a full-time student, or a homemaker. Preinjury competitive employment was also included as a potential predictor.

Injury severity measures included Glasgow Coma Scale motor score at emergency department admission[31]; PTA duration, measured in weeks; and whether the participant was discharged from inpatient rehabilitation in PTA (yes/no). If the participant was discharged in PTA, the total hospital (acute care and inpatient rehabilitation) LOS was used as the PTA duration, thus allowing for fewer missing values and avoiding having the algorithm impute these values. Additional predictors included intracranial pressure elevation (none, <24 hours, >24 hours, sustained for >24 hours, not monitored), craniotomy (yes/no), craniectomy (yes/no), acute hospital LOS, and the presence of a focal traumatic intracranial lesion (subdural or epidural hematoma, or intraparenchymal hemorrhage classified as contusion) found on computed tomographic scan (yes/no).

Statistical methods

The entire sample was split into a training set (85% of the sample) and a test set (15% of the sample). The training set at each follow-up period was used to build the predictive model, whereas the test set was used to assess the predictive ability of the model.

The models were built using the Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE) algorithm.[16] The CRUISE algorithm creates splits based on predictors included in the model and, unlike most other algorithms, allows for multiple, rather than binary, splitting. Variables included at the top of the tree can be considered the most important predictors and variables included lower down in the tree are less related to the outcome. A *node* is a group of similar participants and a *terminal node* is the final node from a set of branches that does not split any further. The algorithm chooses splits to maximize the difference in the outcome using χ^2 tests, so all predictors included in the model are related to the outcome. In addition, linear discriminant analysis is used to select the cut points of continuous measures.[19] Thus, some variables fed into the model may not show up in the final tree if they are not closely related to the outcome. Variables connected down the tree can be grouped together to make predictions.

The CRUISE algorithm options for the current analyses were univariate splitting via the 2-dimension method, with splitting based on discriminant analysis using equal costs of misclassification and equal prior group probabilities. A standard error of 0.05 was used for pruning. To accommodate missing values in predictors, nodewise imputation and fit were employed. A minimum of 150 participants were required in each node to allow a node to split to facilitate a parsimonious model. Finally, a nationally recognized expert panel gave consideration for "manual pruning" of the trees, which involved removal of select lower branches of questionable clinical justification to increase model usability. Manually pruned nodes were unanimously approved by all panel members.

A C-index, based on the percentage of participants with each level of the outcome in each terminal node, was calculated for each tree and used to measure the predictability. The statistic ranges from 0.5 to 1, with 1 indicating perfect predictability and 0.5 meaning random predictability. A value of 0.7 to 0.8 indicates an acceptable model, 0.8 to 0.9 an excellent fit, and greater than 0.9 being an outstanding fit.[32] The index was calculated using the tree based on the training set as well as with the test set.

RESULTS

Descriptive statistics for all eligible participants in the 1-year postinjury sample, as well as for the associated training and test sets, are presented in Table 1. The sample was 75% male and the median injury age was 33.0 years (interquartile range: 23.0-46.0). Although race was not included as a predictor in the model, 68.0% of the sample was white, 17.5% was black, 10.2% was Hispanic, and 4.2% identified as other ethnic groups. Approximately 16.8% of the patients remained in PTA at discharge; among the remainder, the median PTA duration was 4.0 weeks (interquartile range: 2.0-7.0). Overall, the employment rate was 33.6% at 1 year postinjury, increased to 37.5% at 2 years postinjury, and increased to 40.6% at 5 years. Table 2 displays the distribution of employment in the training and test sets.

Each tree was assessed for manual pruning. One branch was pruned from the 1-year postinjury tree and 2 were removed from the 2-year postinjury model. The pruned branches were overly cumbersome, contained small sample sizes, and were inhibiting the interpretability of the trees. The trees prior to manual pruning can be found in the online supplement. No manual pruning was performed in the 5-year postinjury model.

The final decision tree models are provided in Figures 2 to 4 for the 1-, 2-, and 5-year postinjury employment outcomes, respectively. Terminal nodes are distinguished by solid lined boxes. For each model year, PTA duration was the primary split, indicating that it was the most salient predictor of employment. Other variables that appeared frequently in the models were preinjury employment, occupational group, age at injury, and baseline education level. After PTA duration, occupational group and preinjury employment were next in

TABLE 1 *Summary data for predictor candidate variables on those eligible at 1-year postinjury (N = 7867)*[a]

Characteristic	Level	Overall	Training (*N* = 6687)	Test (*N* = 1180)
Preinjury employment	Yes	5069 (73.3%)	4313 (73.4%)	756 (72.7%)
	No	1845 (26.7%)	1561 (26.6%)	284 (27.3%)
Age at injury		33.0 (23.0-46.0)	33.0 (23.0-46.0)	34.0 (23.0-46.0)
Sex	Female	1953 (24.8%)	1673 (25.0%)	280 (23.7%)
	Male	5914 (75.2%)	5014 (75.0%)	900 (76.3%)
Prior TBI	Yes	1411 (17.4%)	1163 (17.4%)	248 (21.0%)
	No	6456 (82.6%)	5524 (82.6%)	932 (79.0%)
Number of prior TBI		0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Preinjury education	<HS/GED	1456 (20.1%)	1255 (20.5%)	201 (18.4%)
	HS/GED	2709 (37.5%)	2272 (37.0%)	437 (40.0%)
	>HS/GED	3062 (42.4%)	2608 (42.5%)	454 (41.6%)
Productivity	Yes	5553 (80.6%)	4711 (80.6%)	842 (81.1%)
	No	1332 (19.4%)	1136 (19.4%)	196 (18.9%)
Occupational category	Professional	1223 (16.2%)	1057 (16.5%)	166 (14.6%)
	Skilled	3101 (41.0%)	2621 (40.8%)	480 (42.4%)
	Manual labor	1418 (18.8%)	1213 (18.9%)	205 (18.1%)
	None	1816 (24.0%)	1534 (23.9%)	282 (24.9%)
Problem alcohol use	Yes	1227 (17.0%)	1043 (17.0%)	184 (17.0%)
	No	5997 (83.0%)	5096 (83.0%)	901 (83.0%)
Illicit drug use	Yes	1801 (23.3%)	1528 (23.3%)	273 (23.6%)
	No	5916 (76.7%)	5034 (76.7%)	882 (76.4%)
PTA duration, wk		4.0 (2.0-6.0)	4.0 (2.0-6.0)	3.0 (1.0-6.0)
Discharged in PTA	Yes	1323 (16.8%)	1133 (16.9%)	190 (16.1%)
	No	6544 (83.2%)	5554 (83.1%)	990 (83.9%)
Baseline motor GCS		6.0 (4.0-7.0)	6.0 (4.0-7.0)	6.0 (4.0-7.0)
Elevated ICP	None	2013 (25.8%)	1719 (25.9%)	294 (25.1%)
	<24 h	874 (11.2%)	757 (11.4%)	117 (10.0%)
	>24 h	985 (12.6%)	850 (12.8%)	135 (11.5%)
	>24 h sustained	204 (2.6%)	175 (2.6%)	29 (2.5%)
	Not monitored	3728 (47.8%)	3130 (47.2%)	598 (51.0%)
Craniotomy	Yes	901 (11.5%)	773 (11.6%)	128 (10.8%)
	No	6966 (88.5%)	5914 (88.4%)	1052 (89.2%)
Craniectomy	Yes	708 (9.0%)	624 (9.3%)	84 (7.1%)
	No	7159 (91.0%)	6063 (90.7%)	1096 (92.9%)
CT focal hemorrhage	Yes	6137 (79.8%)	5203 (79.7%)	934 (80.4%)
	No	1554 (20.2%)	1326 (20.3%)	228 (19.6%)
Acute hospital LOS		18.0 (10.0-28.0)	18.0 (10.0-28.0)	17.0 (10.0-27.0)

Abbreviations: CT, computed tomographic; GCS, Glasgow Coma Scale; ICP, intracranial pressure; LOS, length of stay; PTA, posttraumatic amnesia; TBI, traumatic brain injury.
[a]Continuous variables shown as median (interquartile range); categorical variables shown as *N* (%).

hierarchal importance. The only injury-related variables to emerge in any of the models were acute hospital LOS and presence of craniotomy in the 2-year postinjury trees. Across the years, the participants with shorter PTA duration, who were employed preinjury, and had a higher level of education were more likely to be competitively employed postinjury.

The terminal nodes with the highest and lowest percentages of employed participants are highlighted in Table 3. In all years, the terminal node with the highest employment appears nearly opposite the one with the lowest employment. For example, in year 1, the best terminal node includes those with the shortest PTA and being professionally employed preinjury. This counters the worst node where the participants had the longest PTA and either had nonprofessional employment or were unemployed. At 2 years postinjury, the node with the highest employment rate contains participants with the shortest PTA, greater than high school education, and who were employed preinjury. The worst node includes those with a longer PTA duration and who were not productive preinjury. Five years postinjury, those

TABLE 2 *Distribution of employment and C-statistics by postinjury follow-up year*

Outcome	Summary, *N* (%) Training set	C-statistic (95% CI)	Test set	C-statistic (95% CI)
Employment				
1-y follow-up				
Yes	2248 (33.6%)	0.77 (0.76-0.79)	392 (33.2%)	0.77 (0.74-0.80)
No	4439 (66.4%)		788 (66.8%)	
2-y follow-up				
Yes	2157 (37.3%)	0.77 (0.75-0.78)	389 (38.7%)	0.76 (0.73-0.79)
No	3620 (62.7%)		617 (61.3%)	
5-y follow-up				
Yes	1713 (40.9%)	0.70 (0.69-0.72)	286 (38.7%)	0.72 (0.68-0.76)
No	2475 (59.1%)		453 (61.3%)	

Abbreviation: CI, confidence interval.

with the shorter PTA duration and more education again lead to the best outcome, and the opposite was true for the worst terminal node. These result patterns indicate that the predictors listed previously are positively associated with good outcomes and negatively associated with poor outcomes.

Prediction model performance

The models predicting earlier outcome (eg, 1 year postinjury) had better predictability than long-term (eg 5 years postinjury). The C-statistics ranged from 0.70 to 0.77 (see Table 2), with the 1-year postinjury trees producing the highest C-statistic values indicating better predictability. The second highest was the 2-year postinjury tree and then the 5-year postinjury tree. All models indicate reasonable predictability based on the 0.5 to 1.0 range, and predictability was largely retained in the independent test sets. Thus, the models can be expected to perform well for future patients experiencing a moderate to severe TBI and matriculating to inpatient rehabilitation.

DISCUSSION

Using the TBIMS NDB, the current study utilized classification tree methodology to build practical

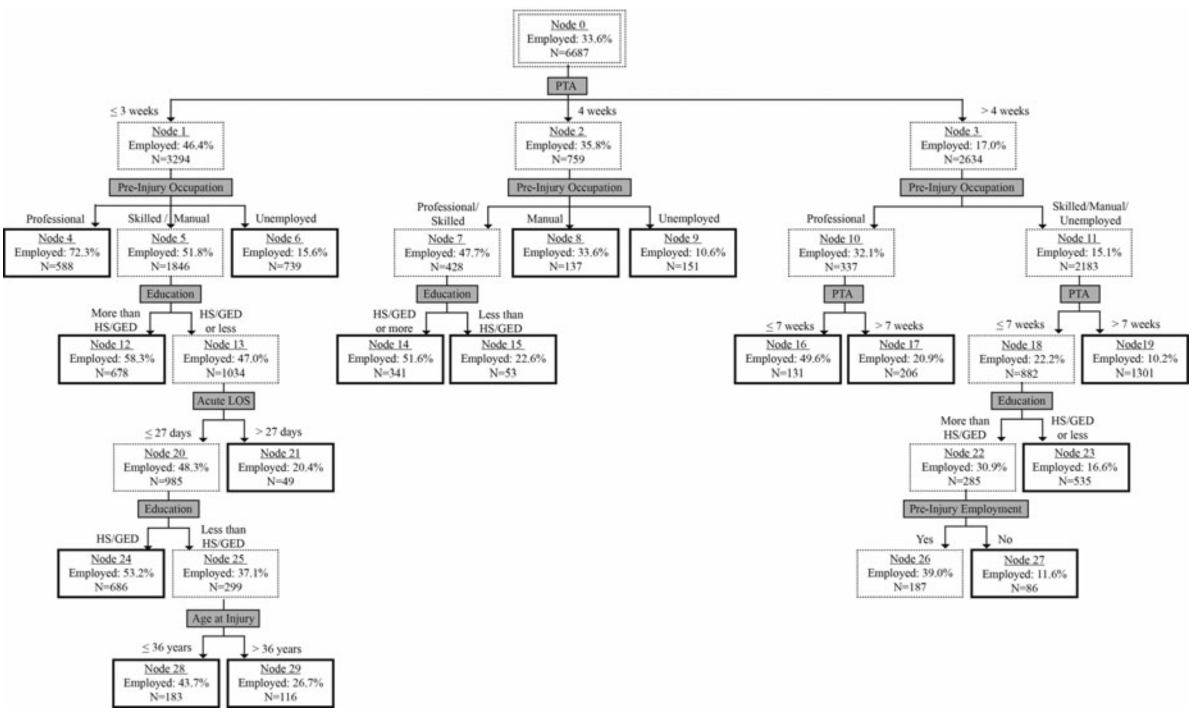

Figure 2. One-year employment classification tree. PTA indicates posttraumatic amnesia.

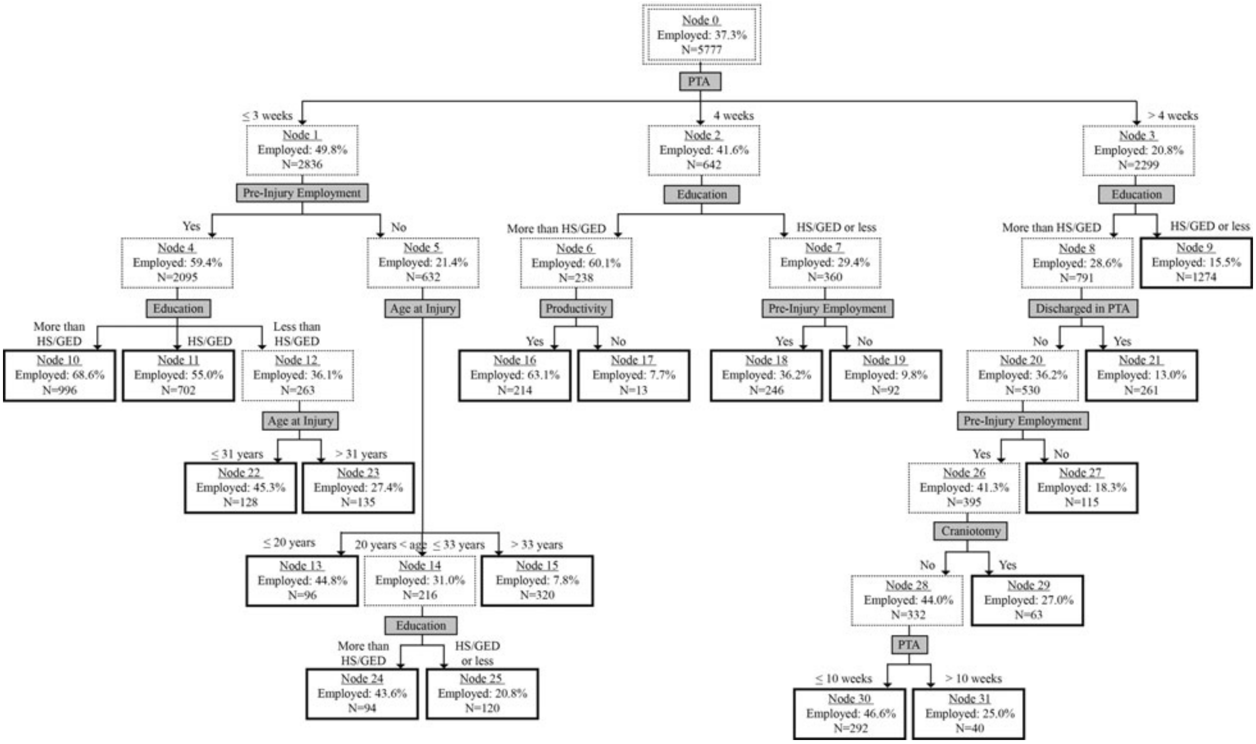

Figure 3. Two-year employment classification tree. PTA indicates posttraumatic amnesia.

prognostic models to characterize employment at 1, 2, and 5 year(s) after moderate to severe closed TBI. Across all 3 follow-up periods, PTA duration emerged as the primary determinant of survivors' likelihood of attaining competitive employment, with longer length of PTA predicting lower chance of postinjury employment. Nearly one-half of all participants who spent less than 3 weeks in PTA were employed at 1 (46.4%) and 2 years (49.8%) postinjury, compared with only 17% and 20.8% who spent greater than 4 weeks in PTA, respectively. Five years after injury, 53.2% of participants who spent at most 4 weeks in PTA were employed compared with 29.4% of participants who were in PTA for longer than 4 weeks (see Figure 2). These findings highlight the 3- to 4-week PTA duration mark as a critical point of demarcation between survivors who have relatively good employment prognoses and those more likely to face significant challenges. The only other injury-related

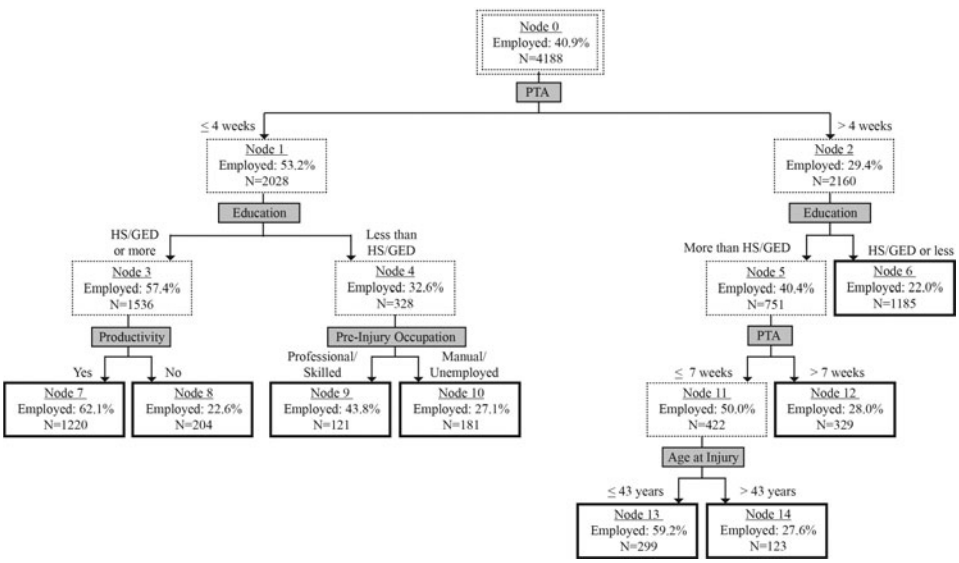

Figure 4. Five-year employment classification tree. PTA indicates posttraumatic amnesia.

TABLE 3 *Terminal nodes showing the highest/lowest prevalence of postinjury employment for each of the 1, 2, and 5-year postinjury follow-up periods*

Postinjury follow-up period	Terminal nodes with highest employment rate				Terminal nodes with lowest employment rate			
	Node	Characteristics	% Employed	% of sample	Node	Characteristics	% Employed	% of sample
1 y	4	PTA ≤3 wk Professionally employed preinjury	72.3	8.8	19	PTA >7 wk Skilled/manual/ unemployed preinjury	10.2	19.5
	12	PTA ≤3 wk Skilled/Manual employment preinjury. More than HS/GED	58.3	10.1	9	PTA = 4 wk Unemployed preinjury	10.6	2.3
	24	PTA ≤3 wk Skilled/Manual employment preinjury HS/GED Acute LOS ≤27 d	53.2	10.3	27	4-7 wk PTA Skilled/manual/ unemployed preinjury. More than HS/GED Not employed preinjury	11.6	1.3
2 y	10	PTA ≤3 wk Employed preinjury More than HS/GED	68.6	17.2	17	PTA = 4 wk More than HS/GED Not productive	7.7	0.2
	16	PTA = 4 wk More than HS/GED Productive	63.1	3.7	15	PTA ≤3 wk Not employed preinjury Age >33 y	7.8	5.5
	11	PTA ≤3 wk Employed preinjury HS/GED	55.0	12.2	19	PTA = 4 wk HS/GED or less Not employed preinjury	9.8	1.6
5 y	7	PTA ≤4 wk HS/GED or more Productive	62.1	29.1	6	PTA >4 wk HS/GED or less	22.0	28.3
	13	4-7 wk PTA More than HS/GED Age ≤43 y	59.2	7.1	8	PTA ≤4 wk HS/GED or more Not productive	22.6	4.9
	9	PTA ≤4 wk Less than HS/GED Professional/Skilled employment pre-injury	43.8	2.9	10	PTA ≤4 wk Less than HS/GED Manual employment/ unemployed preinjury	27.1	4.3

Abbreviations: LOS, length of stay; PTA, posttraumatic amnesia.

characteristics found predictive of employment were acute hospital LOS (year 1), whether discharged in PTA (year 2), and whether a craniotomy was performed (year 2), with all 3 having secondary importance compared with PTA duration. These findings align with our previous research showing that when it comes to predicting long-term functional outcomes after TBI, PTA duration is paramount. It is plausible that the dominance of PTA as a predictor results from PTA subsuming other precursory severity indicators by virtue of conveying more information about rate of recovery.[15] While our method of calculating PTA duration may underestimate the true PTA duration compared with some regression models, 85% of subjects discharged in PTA were done so after 4 weeks postinjury. Thus, even if these subjects had longer PTA than was determined, the vast majority of our sample would remain in the same groupings as we proposed, making it unlikely that the model would yield significantly different results.

After PTA duration, preinjury occupation, employment, and education were the next most important predictors across all models/trees. Our results support previous findings that preinjury employment/vocational status and education are significant predictors of postinjury competitive employment.[3,4,6] These socioeconomic variables may influence employment because they signal innate intelligence, psychosocial characteristics, brain reserve capacity, and/or social support systems.[10,33] Age at injury was the only other variable that appeared in all models/trees, usually emerging lower in the trees than the aforementioned variables.

Whereas our model at 5 years postinjury produced only 2 initial nodes ($\leq$4 weeks and >4 weeks), our models for 1 year and 2 years postinjury each produced 3 PTA nodes ($\leq$3 weeks, 4 weeks, and >4 weeks). The decline of PTA cut points at year 5 may be suggestive of the diminishing importance of the finer gradations of PTA duration for predicting employment over longer postinjury time periods. In other words, PTA duration, and consequently, our predictive model, may lose precision over time, an assertion that is supported by the fact that the 5-year postinjury model had the lowest C-Statistic. Alternately, the decrease in precision could be due to the decrease in the sample size when compared with models derived from the 1- and 2-year follow-up time points.

Other investigators have also reported higher rates of employment over successive years postinjury.[11,34] Our findings are in alignment, with the 5-year postinjury model having the highest overall proportion of employed participants at 40.6%, followed by 37.5% at 2 years and 33.6% at 1 year postinjury. Possible reasons for the improved employment rates over time include decreased focus on medical and therapy appointments, additional brain function recovery, improved adjustment to disability, and later access to vocational rehabilitation resources.[35] Our findings provide additional insight by showing this pattern of better employment rates at 5 years postinjury existed among participants in both the best and worst prognostic nodes. When looking at just the primary splits on days spent in PTA, the best initial node in year 5 (node 1) had a 53.2% employment rate compared with 49.8% (node 1) at 2 years and 46.4% (node 1) at 1 year. Similarly, the worst initial node in each year had employment rates of 17.0% (node 3), 20.8% (node 3), and 29.4% (node 2) for years 1, 2, and 5, respectively (see Figures 2-4).

The finding of improved rate of employment over time in the group with the longest PTA duration, which increased from year 1 (17.0%) to year 2 (20.8%) and to year 5 (29.4%), is promising. Conversely, the finding that 70% of participants who spent more than 4 weeks in PTA remained unemployed at 5 years postinjury is sobering. In sharp contrast, the employment rate among noninstitutionalized individuals in the United States without a disability is 76.4%.[36] The rate of unemployment at 5 years postinjury in our sample (70%) is similar to the rate of unemployment among noninstitutionalized individuals in the United States with a disability (65%). In fact, the rate of employment in the group with the longest duration of PTA at 5 years postinjury in our sample (28.0%) is comparable with the rate of employment among noninstitutionalized individuals in the United States with a cognitive disability (25.6%) but higher than the rates among those with self-care (15.7%) or independent living disability (16.3%). These consistencies support the external validity of our findings.

The decision tree methodology used in this work was inherently data-driven, with no guarantee of consistent patterns. However, we see that similar patterns appear throughout the models. For instance, the decision to split on preinjury occupation status and then on education level presents itself in both nodes 1 and 2 in Figure 2. These discernable patterns are shown in Table 3, such that professional preinjury employment and higher levels of education were associated with greater employment regardless of PTA duration. This trend, and the similar trends in 2 and 5 years postinjury, suggests that while PTA duration is the strongest predictor, the secondary predictors of postinjury employment have a consistent effect.

The current study has some limitations. The decision tree methodology used may yield limited predictive power compared with typical regression modeling,[14] but for clinical applications, the enhanced usability of the resulting model should outweigh this limitation. Branching was limited by sample sizes in terminal nodes getting too small, so some further splitting, with differing predictors, may have been missed.

As with other TBIMS studies, although the NDB is representative of individuals with TBI receiving inpatient rehabilitation,[37,38] findings may not be generalizable to those who do not receive inpatient rehabilitation for their TBI. We must also acknowledge that while several previous studies have found meaningful relationships between race/ethnicity and post-TBI employment,[3,4,9,39] we excluded race/ethnicity from our analyses in efforts to produce parsimonious decision trees with the greatest clinical utility. On the advice of an anonymous reviewer, models with race included were fit for comparison and resulted in worse or comparable predictability at all years, reaffirming our decision for its exclusion. Other variables known to be related to employment, such as preinjury income and mental health problems, were not included because they were unavailable in the data or had uninterpretable definitions in the TBIMS.[25] Finally, we were unable to examine the possible relevance of downstream modifiable factors unknown at baseline, such as psychological symptoms postinjury, substance use, and social support.

CONCLUSION

This is the first study of its kind to use CRUISE decision tree analyses for predicting employment after moderate to severe TBI, and it identified PTA duration, followed by preinjury occupation, employment, and age at injury as the most salient predictors. Although future research is needed to evaluate the external validity of these findings in other cohorts, this study's results have important implications. They shed light on reasonable expectations that can be made near the time of injury about the likelihood of competitive employment at 1, 2, and 5 years postinjury for survivors of moderate-severe closed TBI. There is also potentially valuable information that can be gleaned from these findings about the best candidates for interventions, such as vocational rehabilitation, to maximize employment returns.

REFERENCES

1. Selassie AW, Zaloshnja E, Langlois JA, Miller T, Jones P, Steiner C. Incidence of long-term disability following traumatic brain injury hospitalization, United States, 2003. *J Head Trauma Rehabil*. 2008;23(2):123–131. doi:10.1097/01.HTR.0000314531.30401.39.
2. Prigatano GP, Fordyce DJ, Zeiner HK, Roueche JR, Pepping M, Wood BC. Neuropsychological rehabilitation after closed-head injury in young adults. *J Neurol Neurosurg Psychiatry*. 1984;47(5):505–513. doi:10.1136/jnnp.47.5.505.
3. Gary KW, Ketchum JM, Arango-Lasprilla JC, et al. Differences in employment outcomes 10 years after traumatic brain injury among racial and ethnic minority groups. *J Vocat Rehabil*. 2010;33(1):65–75. doi:10.3233/JVR-2010-0516.
4. Gary KW, Arango-Lasprilla JC, Ketchum JM, et al. Racial differences in employment outcome after traumatic brain injury at 1, 2, and 5 years postinjury. *Arch Phys Med Rehabil*. 2009;90(10):1699–1707. doi:10.1016/j.apmr.2009.04.014.
5. Cuthbert JP, Harrison-Felix C, Corrigan JD, Bell JM, Haarbauer-Krupa JK, Miller AC. Unemployment in the United States after traumatic brain injury for working-age individuals: prevalence and associated factors 2 years postinjury. *J Head Trauma Rehabil*. 2015;30(3):160–174. doi:10.1097/HTR.0000000000000090.
6. Walker WC, Marwitz JH, Kreutzer JS, Hart T, Novack TA. Occupational categories and return to work after traumatic brain injury: a multicenter study. *Arch Phys Med Rehabil*. 2006;87(12):1576–1582. doi:https://doi.org/10.1016/j.apmr.2006.08.335.
7. Conroy L, McKenna K. Vocational outcome following spinal cord injury. *Spinal Cord*. 1999;37(9):624–633.
8. Keyser-Marcus LA, Bricout JC, Wehman P, et al. Acute predictors of return to employment after traumatic brain injury: a longitudinal follow-up. *Arch Phys Med Rehabil*. 2002;83(5):635–641. doi:S0003-9993(02)17128-0 [pii].
9. Arango-Lasprilla JC, Ketchum JM, Williams K, et al. Racial differences in employment outcomes after traumatic brain injury. *Arch Phys Med Rehabil*. 2008;89(5):988–995. doi:10.1016/j.apmr.2008.02.012.
10. Forslund MV, Roe C, Perrin PB, et al. The trajectories of overall disability in the first 5 years after moderate and severe traumatic brain injury. *Brain Inj*. 2017;31(3):329–335. doi:10.1080/02699052.2016.1255778.
11. Kreutzer JS, Marwitz JH, Walker W, et al. Moderating factors in return to work and job stability after traumatic brain injury. *J Head Trauma Rehabil*. 2003;18(2):128–138.
12. Wagner AK, Hammond FM, Sasser HC, Wiercisiewski D. Return to productive activity after traumatic brain injury: relationship with measures of disability, handicap, and community integration. *Arch Phys Med Rehabil*. 2002;83(1):107–114. doi:S0003-9993(02)50008-3 [pii].
13. Temkin NR, Holubkov R, Machamer JE, Winn HR, Dikmen SS. Classification and regression trees (CART) for prediction of function at 1 year following head trauma. *J Neurosurg*. 1995;82(5):764–771. doi:10.3171/jns.1995.82.5.0764 [doi].
14. Rovlias A, Kotsou S. Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *J Neurotrauma*. 2004;21(7):886–893. doi:10.1089/0897715041526249 [doi].
15. Sherer M, Nick TG, Sander AM, et al. Race and productivity outcome after traumatic brain injury: influence of confounding factors. *J Head Trauma Rehabil*. 2003;18(5):408–424.
16. Kim H, Loh WY. Classification trees with unbiased multiway splits. *J Am Stat Assoc*. 2001;96(454):589–604. doi:10.1198/016214501753168271.
17. Fokkema M, Smits N, Kelderman H, Penninx BWJH. Connecting clinical and actuarial prediction with rule-based methods. *Psychol Assess*. 2015;27(2):636–644. doi:10.1037/pas0000072.
18. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009;14(4):323–348. doi:10.1037/a0016973.
19. Hastie T, Tibshirani R, Friedman J. Springer series in statistics the elements of statistical learning the elements of statistical learning. https://web.stanford.edu/~hastie/Papers/ESLII.pdf. Accessed April 30, 2018.
20. Breiman L. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall; 1993.

21. Catalano D, Pereira AP, Wu MY, Ho H, Chan F. Service patterns related to successful employment outcomes of persons with traumatic brain injury in vocational rehabilitation. *Neurorehabilitation*. 2006;21(4):279–293.
22. Kass, GV. An exploratory technique for investigating large quantities of categorical data. *J R Stat Soc C (Applied Stat)*. 1980;29(2):119–127. doi:10.2307/2986296.
23. Harrison-Felix C, Newton CN, Hall KM, Kreutzer JS. Descriptive findings from the Traumatic Brain Injury Model Systems National Data Base. *J Head Trauma Rehabil*. 1996;11:1–14. doi:10.1097/00001199-199610000-00002.
24. Sherer M, Struchen MA, Yablon SA, Wang Y, Nick TG. Comparison of indices of traumatic brain injury severity: Glasgow Coma Scale, length of coma and posttraumatic amnesia. *J Neurol Neurosurg Psychiatry*. 2008;79(6):678–685. doi:10.1136/jnnp.2006.111187.
25. TBINDSC|Syllabus Detail. https://www.tbindsc.org/SyllabusDetail.aspx?MOD=2&ID=EMP. Accessed January 24, 2018.
26. Bogner J, Corrigan JD. Reliability and predictive validity of the Ohio State University TBI identification method with prisoners. *J Head Trauma Rehabil*. 2009;24(4):279–291. doi:10.1097/HTR.0b013e3181a66356.
27. Corrigan JD, Bogner J. Initial reliability and validity of the Ohio State University TBI identification method. *J Head Trauma Rehabil*. 2007;22(6):318–329. doi:10.1097/01.HTR.0000300227.67748.77.
28. Centers for Disease Control and Prevention. The BRFSS data user guide June 2013. https://www.cdc.gov/brfss/data_documentation/pdf/UserguideJune2013.pdf. Published 2013. Accessed January 23, 2018.
29. Centers for Disease Control and Prevention. 2016 Behavioral Risk Factor Surveillance System Questionnaire. https://www.cdc.gov/brfss/questionnaires/pdf-ques/2016_BRFSS_Questionnaire_FINAL.pdf. Published 2016. Accessed December 5, 2017.
30. Sobell LC, Sobell MB. Alcohol consumption measures. https://pubs.niaaa.nih.gov/publications/assessingalcohol/sobell.pdf. Published 2004. Accessed December 5, 2017.
31. Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. *Lancet*. 1974;304(7872):81–84. doi:10.1016/S0140-6736(74)91639-0.
32. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol 398. Hoboken, NJ: John Wiley & Sons; 2013.
33. Levi Y, Rassovsky Y, Agranov E, Sela-Kaufman M, Vakil E. Cognitive reserve components as expressed in traumatic brain injury. *J Int Neuropsychol Soc*. 2013;19(6):664–671. doi:10.1017/S1355617713000192.
34. Forslund MV, Arango-Lasprilla JC, Roe C, Perrin PB, Sigurdardottir S, Andelic N. Multi-level modelling of employment probability trajectories and employment stability at 1, 2 and 5 years after traumatic brain injury. *Brain Inj*. 2014;28(7):980–986. doi:10.3109/02699052.2014.888770.
35. Lukow HR, Godwin EE, Marwitz JH, Mills A, Hsu NH, Kreutzer JS. Relationship between resilience, adjustment, and psychological functioning after traumatic brain injury. *J Head Trauma Rehabil*. 2015;30(4):241–248. doi:10.1097/HTR.0000000000000137.
36. Disability and Employment Status Report for the United States, 2012. http://disabilitystatistics.org/reports/2012/English/HTML/report2012.cfm. Published 2002. Accessed November 26, 2017.
37. Cuthbert JP, Corrigan JD, Whiteneck GG, et al. Extension of the representativeness of the traumatic brain injury model systems national database: 2001 to 2010. *J Head Trauma Rehabil*. 2012;27(6):E15–E27. doi:10.1097/HTR.0b013e31826da983.
38. Corrigan JD, Cuthbert JP, Whiteneck GG, et al. Representativeness of the traumatic brain injury model systems national database. *J Head Trauma Rehabil*. 2012;27(6):391–403. doi:10.1097/HTR.0b013e3182238cdd.
39. Arango-Lasprilla JC, Ketchum JM, Gary KW, et al. The influence of minority status on job stability after traumatic brain injury. *PMR*. 2009;1(1):41–49. doi:10.1016/j.pmrj.2008.07.001.