

SUPPLEMENTAL MATERIAL

Expanded Methods

1. Summative measure of medical conditions and health-risk behaviors

The medical conditions ascertained from medical records included Alzheimer's disease or dementia, coronary artery disease or myocardial infarction, atrial fibrillation, heart failure, cancer, chronic obstructive pulmonary disease, diabetes, end-stage renal disease, epilepsy, high cholesterol, hypertension, Parkinson's disease.

2. Multiple imputation

PROC MI was used to create imputed data sets. The FCS statement was used for fully conditional specification methods, because the missing data pattern was arbitrary and imputed variables include both categorical and continuous variables. More specifically, the logistic regression method was used for binary or ordinal variables (education, stroke type, the categorized Modified Rankin Scale (MRS), pre-stroke depression, current smoking, obesity). The predictive mean matching method was used for continuous variables (scores of the National Institutes of Health Stroke Scale (NIHSS), and scores of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)). The number of imputations was set to ten, and the number of burn-in iterations before each imputation was set to 100. Variables used for multiple imputation models included age, sex, race-ethnicity, education, marital status, insurance status, stroke type and stroke severity, medical conditions, health-risk behaviors, functional limitation, cognitive function and pre-stroke depression. PROC MIANALYZE was used to combine the results of the imputation-based analyses and make statistical inferences.

3. Inverse probability weighting

To determine predictors for generating the inverse probability weights, we fitted three prediction models using complete case analysis. Characteristics of the three samples are presented in Supplemental Tables S1-S3. Functional forms of continuous variables were determined using the LOWESS (locally weighted scatterplot smoothing) smoothed logit. Linear and squared terms of the same variable were grouped together in the backward-selection search (i.e. appeared or not appeared together). Predictors for

each model were selected based on backward stepwise logistic regression with a significance level of 0.20 for removal from the model (Supplemental Table S4).

In the first prediction model (Prediction Model 1), the probability of participating in the interview portion of the BASIC study among patients who survived 90 days after stroke was modeled as a function of race-ethnicity, insurance status, stroke severity (log-transformed NIHSS score and log-transformed NIHSS score squared) and history of congestive heart failure.

In the second prediction model (Prediction Model 2), the probability of having a non-proxy baseline interview (versus proxy interview) among participants who completed the baseline interview was modeled as a function of age, insurance status, stroke severity (log-transformed NIHSS score and log-transformed NIHSS score squared), stroke type, baseline functional limitations (MRS scores), pre-stroke cognitive function (IQCODE scores), medical conditions (Alzheimer's disease or dementia, congestive heart failure, atrial fibrillation and cancer) and excessive alcohol use.

In the third prediction model (Prediction Model 3), the probability of completing the outcome interview at 90 days among participants with non-proxy baseline interview was modeled as a function of age, sex, education, stroke severity (log-transformed NIHSS scores and log-transformed NIHSS score squared), and pre-stroke functional limitation.

Supplemental Table S1. Sample Characteristics of the Prediction Model 1

	Total (n=1360)	Patients who participated in the baseline interview (n=1034)	Patients who did not participate in the baseline interview (n=326)	<i>P</i> value
Age, mean (SD)	67.8 (12.3)	67.6 (12.1)	68.4 (12.9)	0.367
Sex, N (%)				0.709
Male	722 (53.1)	546 (52.8)	176 (54.0)	
Female	638 (46.9)	488 (47.2)	150 (46.0)	
Ethnicity, N (%)				0.008
Non-Hispanic Whites	545 (40.1)	394 (38.1)	151 (46.3)	
Mexican Americans	815 (59.9)	640 (61.9)	175 (53.7)	
Insurance status, N (%)				0.210
Uninsured	165 (12.1)	119 (11.5)	46 (14.1)	
Insured	1195 (87.9)	915 (88.5)	280 (85.9)	
Stroke type*, N (%)				0.271
Ischemic stroke	1190 (87.8)	900 (87.2)	290 (89.5)	
Intracerebral hemorrhage stroke	166 (12.2)	132 (12.8)	34 (10.5)	
Stroke severity (NIHSS)*, mean (SD)	5.5 (6.5)	5.6 (6.5)	5.1 (6.5)	<0.01
Medical conditions, N (column %)				
Dementia/Alzheimer's disease	85 (6.3)	67 (6.5)	18 (5.5)	0.533
Atrial fibrillation	163 (12.0)	121 (11.7)	42 (12.9)	0.567
Congestive heart failure	95 (7.0)	64 (6.2)	31 (9.5)	0.040
Coronary artery disease	334 (24.6)	255 (24.7)	79 (24.2)	0.875
Hypertension	1056 (77.7)	808 (78.1)	248 (76.1)	0.434
Diabetes	609 (44.8)	464 (44.9)	145 (44.5)	0.900
Cancer	156 (11.5)	120 (11.6)	36 (11.0)	0.781
COPD	116 (8.5)	84 (8.1)	32 (9.8)	0.340
Health-risk behaviors, N (column %)				
Current smoking*	316 (24.1)	236 (23.7)	80 (25.2)	0.588
Excessive alcohol use	127 (9.3)	95 (9.2)	32 (9.8)	0.734
Obesity*, N (column %)	513 (37.8)	397 (38.5)	116 (35.7)	0.362

Abbreviations: COPD, chronic obstructive pulmonary disease; NIHSS, the National Institutes of Health Stroke Scale; SD, standard deviation.

*Variables with missing data. The numbers of missing values are 4 for stroke type, 5 for NIHSS, 49 for current smoking, and 4 for obesity.

Supplemental Table S2. Sample Characteristics of the Prediction Model 2

	Total (n=1034)	Patients who had non-proxy baseline interview (n=801)	Patients who had proxy baseline interview (n=233)	<i>P</i> value
Age, mean (SD)	67.6 (12.1)	65.7 (11.2)	74.1 (12.7)	<0.001
Sex, N (%)				0.052
Male	546 (52.8)	436 (54.4)	110 (47.2)	
Female	488 (47.2)	365 (45.6)	123 (52.8)	
Ethnicity, N (%)				0.035
Non-Hispanic Whites	394 (38.1)	319 (39.8)	75 (32.2)	
Mexican Americans	640 (61.9)	482 (60.2)	158 (67.8)	
Education*, N (%)				<0.001
Below high school	346 (33.6)	243 (30.3)	103 (45.0)	
High school	288 (28.0)	235 (29.3)	53 (23.1)	
Vocational/some college	256 (24.9)	215 (26.8)	41 (17.9)	
College or more	140 (13.6)	108 (13.5)	32 (14.0)	
Marital status, N (%)				0.811
Married or living together	512 (49.6)	399 (49.8)	113 (48.9)	
Single/widowed/divorced /Separated	520 (50.4)	402 (50.2)	118 (51.1)	
Insurance status, N (%)				<0.001
Uninsured	119 (11.5)	108 (13.5)	11 (4.7)	
Insured	915 (88.5)	693 (86.5)	222 (95.3)	
Stroke type*, N (%)				<0.001
Ischemic stroke	900 (87.2)	727 (91.0)	173 (74.3)	
Intracerebral hemorrhage stroke	132 (12.8)	72 (9.0)	60 (25.7)	
Stroke severity (NIHSS)*, mean (SD)	5.6 (6.5)	4.2 (4.9)	10.5 (8.7)	<0.001
Cognitive function (IQCODE)*, mean (SD)	3.2 (0.4)	3.1 (0.3)	3.3 (0.5)	<0.001
Modified Rankin Scale*, N (%)				<0.001
0-1	472 (46.4)	408 (51.7)	64 (28.2)	
2-3	460 (45.2)	346 (43.8)	114 (50.2)	
4-5	85 (8.4)	36 (4.5)	49 (21.6)	
Medical conditions, N (column %)				
Dementia/Alzheimer's disease	67 (6.5)	23 (2.9)	44 (18.9)	<0.001
Atrial fibrillation	121 (11.7)	88 (11.0)	33 (14.2)	0.184
Congestive heart failure	64 (6.2)	41 (5.1)	23 (9.8)	0.008
Coronary artery disease	255 (24.7)	188 (23.5)	67 (28.8)	0.100

Hypertension	808 (78.1)	615 (76.8)	193 (82.8)	0.049
Diabetes	464 (44.9)	365 (45.6)	99 (42.5)	0.406
Cancer	120 (11.6)	98 (12.2)	22 (9.4)	0.241
COPD	84 (8.1)	65 (8.1)	19 (8.2)	0.984
Health-risk behaviors, N (column %)				
Current smoking*	236 (23.7)	197 (25.6)	39 (17.5)	0.013
Excessive alcohol use	95 (9.2)	73 (9.1)	22 (9.4)	0.879
Obesity*, N (column %)	397 (38.5)	320 (40.0)	77 (33.3)	0.067

Abbreviations: COPD, chronic obstructive pulmonary disease; IQCODE, the Informant Questionnaire on Cognitive Decline in the Elderly; NIHSS, the National Institutes of Health Stroke Scale; SD, standard deviation.

*Variables with missing data. The numbers of missing values are 4 for education, 2 for stroke type, 5 for NIHSS, 17 for Modified Rankin Scale, 136 for IQCODE, 40 for current smoking, and 3 for obesity.

Supplemental Table S3. Sample Characteristics of the Prediction Model 3

	Total (n=801)	Patients who participate in the outcome interview (n=616)	Patients who did not participate in the outcome interview (n=185)	<i>P</i> value
Age, mean (SD)	65.7 (11.2)	66.2 (11.3)	63.9 (10.9)	0.015
Sex, N (%)				0.002
Male	436 (54.4)	317 (51.5)	119 (64.3)	
Female	365 (45.6)	299 (48.5)	66 (35.7)	
Ethnicity, N (%)				0.908
Non-Hispanic Whites	319 (39.8)	246 (39.9)	73 (39.5)	
Mexican Americans	482 (60.2)	370 (60.1)	112 (60.5)	
Education, N (%)				0.151
Below high school	243 (30.3)	179 (29.1)	64 (34.6)	
High school	235 (29.3)	173 (28.1)	62 (33.5)	
Vocational/some college	215 (26.8)	177 (28.7)	38 (20.5)	
College or more	108 (13.5)	87 (14.1)	21 (11.4)	
Marital status, N (%)				0.388
Married or living together	399 (49.8)	312 (50.7)	87 (47.0)	
Single/widowed/divorced or separated	402 (50.2)	304 (49.3)	98 (53.0)	
Insurance status, N (%)				0.048
Uninsured	108 (13.5)	75 (12.2)	33 (17.8)	
Insured	693 (86.5)	541 (87.8)	152 (82.2)	
Stroke type*, N (%)				0.316
Ischemic stroke	727 (91.0)	563 (91.5)	164 (89.1)	
Intracerebral hemorrhage stroke	72 (9.0)	52 (8.5)	20 (10.9)	
Stroke severity (NIHSS)*, mean (SD)	4.2 (4.9)	4.1 (4.7)	4.6 (5.6)	0.517
Cognitive function (IQCODE)*, mean (SD)	3.1 (0.3)	3.1 (0.3)	3.1 (0.3)	0.296
Functional limitation (MRS)*, N (%)				0.844
No symptoms or disability (0-1)	408 (51.7)	314 (51.7)	94 (51.4)	
Slight/moderate disability (2-3)	346 (43.8)	264 (43.5)	82 (44.8)	
Moderately severe/severe disability (4-5)	36 (4.5)	29 (4.8)	7 (3.8)	
Medical conditions, N (%)				
Dementia/Alzheimer's disease	23 (2.9)	21 (3.4)	2 (1.1)	0.096
Atrial fibrillation	88 (11.0)	69 (11.2)	19 (10.3)	0.723
Congestive heart failure	41 (5.1)	31 (5.0)	10 (5.4)	0.840
Coronary artery disease	188 (23.5)	153 (24.8)	35 (18.9)	0.096

Hypertension	615 (76.8)	476 (77.3)	139 (75.1)	0.546
Diabetes	365 (45.6)	275 (44.6)	90 (48.7)	0.337
Cancer	98 (12.2)	76 (12.3)	22 (11.9)	0.871
COPD				0.355
Health-risk behaviors, N (%)				
Current smoking*	197 (25.6)	148 (25.1)	49 (27.1)	0.592
Excessive alcohol use	73 (9.1)	51 (8.3)	22 (11.9)	0.134
Obesity*	320 (40.0)	252 (40.9)	68 (37.0)	0.337
Pre-stroke depression*, N (%)				0.395
None	532 (67.3)	403 (66.4)	129 (70.5)	
Past depression history	129 (16.3)	99 (16.3)	30 (16.4)	
Current antidepressant use	129 (16.3)	105 (17.3)	24 (13.1)	

Abbreviations: COPD, chronic obstructive pulmonary disease; IQCODE, the Informant Questionnaire on Cognitive Decline in the Elderly; MRS, modified Rankin scale; NIHSS, the National Institutes of Health Stroke Scale; SD, standard deviation.

*Variables with missing data. The numbers of missing values are 2 for stroke type, 3 for NIHSS, 11 for Modified Rankin Scale, 133 for IQCODE, 30 for current smoking, 1 for obesity, and 11 for history of depression and antidepressant use.

Supplemental Table S4. Prediction Models Based on Backward Stepwise Logistic Regression

	Prediction Model 1 (n=1298)	Prediction Model 2 (n=842)	Prediction Model 3 (n=623)
Age		0.95 (0.94, 0.97)	1.02 (1.00, 1.04)
Sex			
Male			Reference
Female			1.89 (1.24, 2.87)
Ethnicity			
Non-Hispanic Whites	Reference		
Mexican Americans	1.43 (1.10, 1.85)		
Insurance status			
Insured	Reference		
Uninsured	0.75 (0.52, 1.10)		
Stroke type			
Ischemic stroke		Reference	
Intracerebral hemorrhage stroke		0.38 (0.23, 0.62)	
Stroke severity (NIHSS)			
Log NIHSS	1.51 (1.01, 2.26)	1.55 (0.80, 2.98)	1.45 (0.76, 2.77)
Log NIHSS squared	0.91 (0.80, 1.04)	0.67 (0.55, 0.82)	0.81 (0.66, 1.01)
Cognitive function (IQCODE)			
Normal		Reference	
CIND		1.15 (0.74, 1.78)	
Dementia		0.57 (0.32, 1.03)	
Pre-stroke disability(MRS)			
No symptoms or disability (0-1)		Reference	
Slight/moderate disability (2-3)		0.62 (0.40, 0.94)	
Moderately severe/severe disability (4-5)		0.27 (0.14, 0.54)	
Medical conditions			
Dementia/Alzheimer's disease		0.29 (0.14, 0.62)	
Congestive heart failure	0.51 (0.32, 0.82)	0.57 (0.27, 1.20)	
Atrial fibrillation		1.80 (0.96, 3.40)	
Cancer		2.76 (1.39, 5.49)	
Excessive alcohol use		0.61 (0.32, 1.14)	
Obesity			1.33 (0.87, 2.04)

Abbreviations: CIND, cognitive impairment no dementia; IQCODE, the Informant Questionnaire on Cognitive Decline in the Elderly; MRS, modified Rankin scale; NIHSS, the National Institutes of Health Stroke Scale.

Supplemental Table S5. Results from the fully adjusted multivariable logistic regression of the association between ethnicity and prevalence of post-stroke depression adjusted for selected covariates, the Brain Attack Surveillance in Corpus Christi study, United States, 2011-2015.

	Adjusted for attrition (n=586) OR (95% CI)
Ethnicity	
Mexican Americans	1.15 (0.68, 1.95)
Non-Hispanic Whites	Reference
Age	0.97 (0.95, 0.99)
Sex	
Male	Reference
Female	0.99 (0.62, 1.59)
Education	
Below high school	1.55 (0.84, 2.85)
High school	Reference
Vocational/some college	0.94 (0.52, 1.70)
College or more	0.61 (0.26, 1.42)
Insurance status	
Insured	Reference
Uninsured	1.24 (0.63, 2.44)
Marital status	
Married or living together	Reference
Single/separated/divorced	1.20 (0.76, 1.90)
Stroke type	
Ischemic stroke	Reference
Intracerebral hemorrhage stroke	0.97 (0.45, 2.10)
Stroke severity (log-transformed NIHSS)	1.16 (0.90, 1.50)
Pre-stroke disability (MRS)	
No symptoms or disability (0-1)	Reference
Slight/moderate disability (2-3)	1.39 (0.85, 2.27)
Moderately severe/severe disability (4-5)	1.89 (0.72, 4.98)
Cognitive function (IQCODE)	
Normal (0-3)	Reference
CIND (3.01-3.43)	0.98 (0.56, 1.69)
Dementia (≥ 3.44)	1.16 (0.53, 2.49)
Pre-stroke depression	
None	Reference
Past depression history	2.33 (1.29, 4.22)
Antidepressant use at stroke onset	4.94 (2.74, 8.92)
Number of medical conditions	1.15 (0.99, 1.33)
Current smoking	
No	Reference
Yes	2.38 (1.43, 3.97)
Excessive alcohol use	
No	Reference
Yes	0.96 (0.43, 2.13)
Obesity	
No	Reference
Yes	1.34 (0.84, 2.13)