Paper 1494-2014

Using SAS/STAT® Software to Validate a Health Literacy Prediction Model in a Primary Care Setting

Lucy D'Agostino McGowan, MS, Washington University School of Medicine;
Melody S. Goodman, PhD, Washington University School of Medicine;
Kimberly A. Kaphingst, Washington University School of Medicine

ABSTRACT

Existing health literacy assessment tools developed for research purposes have constraints that limit their feasibility for use in clinical settings. The measurement of health literacy in clinical practice can be impractical due to the staff burden and time requirements of existing assessment tools. Single Item Literacy Screener (SILS) items, which are self-administered brief screening questions, have been developed to address this constraint. We developed a model to predict limited health literacy that consists of two SILS and demographic information (age, gender, race, and education) using a sample of patients in a St. Louis emergency department. In this paper, we validate this prediction model in a separate sample of patients visiting a safety net primary care clinic in St. Louis. Using the prediction model developed in the previous study, we use SAS/STAT® software to validate this model based on two goodness of fit criteria: rescaled R-squared and AIC. The Rapid Assessment of Health Literacy in Medicine – Revised (REALM-R) is used as the gold standard health literacy measure. We evaluate the prediction model by examining the concordance, area under the ROC curve, sensitivity, specificity, and kappa statistics. Using the Youden Index, we choose a cutpoint of 0.54 to define health literacy status. Results show 73.0% concordance when comparing the model estimation to the REALM-R. Development of a validated prediction model for inadequate health literacy based on self-reported data would provide a feasible way to assess health literacy in fast-paced clinical settings. This would allow for the identification of patients with limited health literacy and the implementation of personalized care and communication interventions (e.g., teach-back) that meet their information needs.

INTRODUCTION

Health literacy is defined by the CDC as the "capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions".[1] Several tools have been developed to assess health literacy; however, existing health literacy assessment tools developed for research purposes have constraints that limit their utility in face-paced clinical settings, including requirements for staff time and for verbal administration.

The Short Test of Functional Health Literacy in Adults (S-TOFHLA) requires up to seven minutes for completion and verbal administration.[2–4] The Newest Vital Sign (NVS) requires between two and three minutes to complete and verbal administration.[5] While the Rapid Assessment of Health Literacy in Medicine – Revised (REALM-R) takes less than two minutes, it also requires verbal administration. Regardless of the time component, the verbal administration requirement can be less feasible in a clinical environment, particularly in safety net primary care settings.

Single Item Literacy Screener (SILS) items, on the other hand, are self-administered. The accuracy of the SILS for measuring health literacy in comparison to the REALM has previously been reported.[6,7] We hypothesize that a combination of demographic characteristics (age, gender, race, education) known to be associated with health literacy, in combination with the SILS, will improve our ability to predict limited health literacy using self-reported data.

METHODS

We developed an original prediction model based on data collected from a cross-sectional convenience sample of 425 patients that were recruited from an urban academic emergency department in St. Louis. Two prediction models were developed; one using the Rapid Assessment of Adult Literacy in Medicine – Revised (REALM-R) as the health literacy gold standard and the other using the abbreviated Short Test of Functional Health Literacy in Adults (S-TOFHLA) as the criterion standard. The best prediction model for both outcomes included: age, gender, race, education attainment, and two SILS items. Here we validate the model developed using the emergency department sample with the REALM-R as the criterion standard in a sample of 408 patients visiting a safety net primary care clinic in St. Louis that serves a similar patient population.

Health Literacy Criterion Standard

The Rapid Assessment of Health Literacy in Medicine – Revised (REALM-R) is a health literacy assessment (word recognition test) in which participants are asked to pronounce 11 common medical terms: "fat," "flu," "pill," "allergic," "jaundice," "anemia," "fatigue," "directed," "colitis," "constipation," and "osteoporosis." The first three words are not counted as part of the REALM-R score, but are included to decrease test anxiety. A trained REALM-R administrator scores the pronunciation (correct/incorrect) of each of the other eight words.[8] The assessment is scored by giving one point for each word pronounced correctly, resulting in 8 possible points. We dichotomized the REALM-R score into inadequate health literacy (scores 0-6) and adequate health literacy (scores >6) using standard scoring.

The REALM-R was derived in 157 primarily Caucasian, well-educated adults in the Internal Medicine Clinic at the University of Kentucky. The instrument was validated against the full REALM (66 words) in a distinct population of 203 patients from four university hospital clinics as well as, among one hundred state prison inmates, against the Slosson Oral Reading Test (SORT), Peabody Individual Achievement Test-Revised (PIAT-R), and the Wide Range Achievement Test-Revised (WRAT-R) with correlations of 0.97, 0.96, and 0.88, respectively.[9,10] The REALM-R was correlated with the WRAT-R with a Spearman Rank Correlation of 0.64.[6,8] The Cronbach's alpha between the REALM-R and the WRAT-R was 0.91.[8]

Single Item Literacy Screener (SILS) Items

Participants were also asked three Single Item Literacy Screener (SILS) items[11,12].

1. "How often do you have problems learning about your medical condition because of difficulty understanding written information?" (1="always", 2="often", 3="sometimes", 4="rarely", 5= "never").

2. "How confident are you filling out medical forms by yourself?" (1="not at all", 2="a little bit", 3="somewhat", 4="quite a bit", 5="extremely" confident)

3. "How often do you have someone help you read hospital materials?" (1="always", 2="often", 3="sometimes", 4="rarely", 5= "never")

Prior validation studies show that the screener has strong psychometric properties and is highly correlated with inadequate health literacy as assessed by the Short Test of Functional Health Literacy in Adults (STOFHLA)[3,6,13] and the Rapid Estimate of Adult Literacy in Medicine (REALM).[6,9,14] As previously stated in the literature, scores less than four on each of the SILS were considered to indicate that the participant had difficulty with the material in question.[15]

DATA ANALYSIS

Data analysis was conducted using SAS/STAT® 9.4; statistical significance assessed as P<.05. We began with a base multivariable logistic regression model (PROC logistic) that included the following demographic information: age (modeled continuously), gender (dichotomous; female/male) race (dichotomous; Black/White), and education (categorical; less than high school/high school diploma/more than high school). We next tested each dichotomous SILS item (<4/≥4) individually, by adding them to the base model one at a time. We then looked at all three SILS items in the model. We ultimately chose a model that has the best goodness of fit as determined by the maximum rescaled R-squared and the Akaike Information Criteria (AIC). For R-squared, values closer to 1 indicate better prediction and for the AIC smaller values indicate better prediction. The following code was used:

```
proc logistic data=model PLOTS(ONLY)=(ROC(ID=prob));
      class var1 var2 var3 …/param=ref;
      model REALM= var1 var2 var3 …/outroc=rocout_items ctable rsq;
      roc;
      output out=estimated predicted=estprob l=lower95 u=upper95;
run;
```

In the code above, `PLOTS(ONLY)` option allows only the requested plots to be displayed. In this example, only the ROC plot will be in the output. The option `ID=prob` within the ROC plot option requests that the predicted probabilities are displayed on the ROC plot (Figure 1). The categorical variables should be listed in the `class` statement. The `outroc=` option outputs a dataset that contains the data used to produce the ROC curve. The `ctable` option requests a classification table to be output. The `roc` statement will provide information about the ROC curve, including the area under the curve. The `output` statement will request an output dataset to be created.

Using the final prediction model, we estimated the probability of inadequate health literacy for each participant. Using the receiver operating characteristic (ROC) curve, as well as the sensitivity and specificity table, we determined an inadequate health literacy cutoff. We used the area under the ROC curve, sensitivity, specificity, and the kappa statistic to determine the quality of the prediction model. The area under the ROC curve ranges from 0.5 for a model with no discriminatory ability to 1.0 for a model that is perfectly accurate. The kappa statistic is a measure of inter-rater agreement that takes into account the agreement that is possible by chance. This kappa statistic is calculated from the following formula:

$$\kappa = \frac{P_O - P_e}{1 - P_e}$$

Where P_O is the observed agreement between the two assessments and P_e is the hypothetical probability of agreement by chance.[16]

The simple kappa coefficient ranges from -1 to 1 with values less than 0 indicating that the observed agreement is less likely than the chance agreement, a kappa value of 0 indicating that the observed agreement is as likely as an agreement by chance, and a positive kappa value indicating that the agreement is more likely than that by chance.

In order to examine the concordance, discordance, and the kappa statistic, the following code is used:

```
proc freq data=data;
table REALM*model_prediction/agree;
run;
```

In the code above, the `agree` statement requests agreement statistics, such as the kappa statistic, to be calculated. Concordance is calculated as the percent that have the same health literacy status as assessed by both the REALM-R and the prediction model. Discordance is calculated as 1-concordance.

We then used PROC CORR to examine the correlation between the predicted outcome and the REALM-R determined outcome using the Pearson Correlation Coefficient and the reliability of the prediction using Cronbach's alpha. Cronbach's alpha is often used as a measure of internal consistency to determine validity. In general, an alpha value of 0.7 is acceptable for validation.[17,18]

```
proc corr data=data alpha nomiss;
var REALM model_prediction;
run;
```

Here, the `alpha` option requests the Cronbach Coefficient Alpha to be reported and the `nomiss` ensures that observations with missing variables are deleted in a listwise manner in order to have the same number of observations across all variables included in the calculation.[19]

We validate the health literacy prediction model in the primary care sample by examining the concordance and discordance rates, the kappa statistic, correlation, and Cronbach's alpha; the REALM-R is the criterion standard.

RESULTS

After testing the six potential models the final model was selected based on goodness of fit criterion: R-squared and AIC (Table 1). Based on the emergency department sample, the probability of inadequate health literacy for each patient was estimated with the following model:

$$\begin{aligned} Y = -0.3676 - \ 0.0139(age) - \ 0.5069(female) + \ 1.7075(Black\ race) + \ 0.7262(< 12\ grade\ education) \\ - \ 1.5114(> 12\ grade\ education) + \ 1.7860\ \big(ability\ to\ read(inadequate)\big) \\ + 0.8822\ (help\ reading\ hospital\ materials(inadequate)) \end{aligned}$$

Where age is modeled continuously, gender is dichotomous with male the reference, race is dichotomous with White as the reference, education is categorical with high school diploma as the reference, and the SILS items are dichotomized at 4 on the Likert scale, comparing participants that responded that they "rarely/never" (reference group) require help to those that responded "always/often/sometimes". For both SILS items included in the model, "ability to read" and "help reading hospital materials" the Likert scale options are: 1="always", 2="often", 3="sometimes", 4="rarely", 5= "never".

Table 1. Health Literacy Prediction Model Comparisons

The table below shows the comparison of the models to predict inadequate health literacy. Model 6 was chosen as the final model based on the model goodness of fit statistics, maximum rescaled R-squared (R^2) and Akaike information criterion (AIC).

	Predictors								Model Goodness of Fit Statistics	
Models	**Age**	**Gender (ref: Male)**	**Race (ref: White)**	**Education (ref: High School)**		**Ability to read (ref: need help rarely/ never)**	**Confidence filling out Medical forms (ref: quite a bit/ extremely confident)**	**Help reading hospital materials (ref: need help rarely/ never)**	**R^2**	**AIC**
		Female	Black	<High School	>High school	Need help always/ often/ sometimes	not at all/ a little bit/ somewhat confident	Need help always/ often/ sometimes		
Model 1									0.38	462.4
OR	0.99	0.58	6.06	3.05	0.21	-	-	-		
95% CI	(0.97, 1.00)	(0.36, 0.94)	(3.58, 19.28)	(1.55, 6.00)	(0.13, 0.36)	-	-	-		
p-value	0.130	0.026	<0.001	0.001	<0.001	-	-	-		
Model 2									0.43	441.5
OR	0.99	0.58	5.38	2.28	0.22	8.15	-	-		
95% CI	(0.97, 1.00)	(0.35, 0.94)	(3.13, 9.25)	(1.12, 4.66)	(0.13, 0.38)	(2.97, 22.35)	-	-		
p-value	0.053	0.028	<0.001	0.024	<0.001	<0.001	-	-		
Model 3									0.40	455.0
OR	0.99	0.62	5.88	2.55	0.22	-	2.27	-		
95% CI	(0.97, 1.00)	(0.38, 1.00)	(3.44, 10.03)	(1.27, 5.11)	(0.13, 0.37)	-	(1.34, 3.84)	-		
p-value	0.132	0.051	<0.001	0.009	<0.001	-	0.002	-		
Model 4									0.41	446.6
OR	0.99	0.61	6.04	2.50	0.22	-	-	3.25		
95% CI	(0.97, 1.00)	(0.38, 1.00)	(3.51, 10.39)	(1.24, 5.04)	(0.13, 0.37)	-	-	(1.76, 6.01)		
p-value	0.152	0.048	<0.001	0.011	<0.001	-	-	<0.001		
Model 5									0.44	435.7
OR	0.99	0.61	5.52	2.04	0.22	5.44	1.22	2.23		
95% CI	(0.97, 1.00)	(0.37, 1.00)	(3.17, 9.59)	(0.98, 4.24)	(0.13, 0.39)	(1.86, 15.87)	(0.66, 2.26)	(1.12, 4.45)		
p-value	0.090	0.050	<0.001	0.055	<0.001	0.002	0.517	0.023		
Model 6									0.44	434.2
OR	0.99	0.60	5.52	2.07	0.22	5.97	-	2.42		
95% CI	(0.97, 1.00)	(0.37, 0.99)	(3.18, 9.58)	(1.00, 4.29)	(0.13, 0.38)	(2.11, 16.85)	-	(1.26, 4.63)		
p-value	0.082	0.044	<0.001	0.051	<0.001	<0.001	-	0.008		

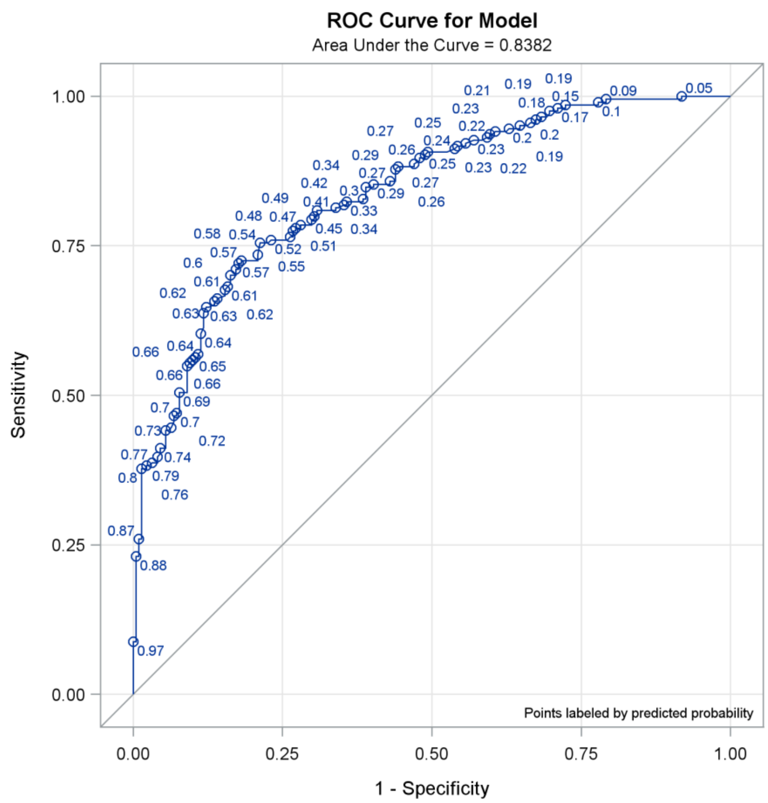
Figure 1 shows the ROC curve comparing the health literacy prediction model (Model 6 in Table 1) to REALM-R. The area under the ROC curve is 0.84 (95% CI: 0.80, 0.88), suggesting that this model predicts inadequate health literacy very well. The Youden Index is frequently used to choose a cutoff point.[20] This index is defined as:

$$J = maximum\{Sensitivity(c) + Specificity(c) - 1\}$$

Where c is each possible cutpoint and J, occurring at the optimal cutpoint, maximizes the difference between the ROC curve and the vertical line of chance.[21]

Using the ROC curve and the Youden Index, we chose 0.54 as the cutoff value. This cutoff yields a sensitivity of 75.5% and a specificity of 78.3%. Within the emergency department sample, we compared the final prediction model using the 0.54 cutoff for inadequate health literacy to the REALM-R, yielding a 76.9% concordance and a kappa statistic of 0.54 (95% CI: 0.46, 0.62). The model-predicted health literacy status was correlated with the REALM-R prediction with a Pearson Correlation Coefficient of 0.54 ($p<0.001$) and a Cronbach alpha of 0.70. While this alpha is considered marginally acceptable, it is important to note we are only comparing two items; alpha is dependent on the number of items being measured and could be low due to a small number of items in the analysis.[17]

Figure 1. The ROC curve comparing the REALM-R to the health literacy prediction model in the emergency department sample



Validation in the Primary Care Sample

In order to validate the health literacy prediction model developed in the emergency department sample, we tested the concordance and discordance in the primary care sample. We found that the concordance was 73.0% (Table 2), with a kappa statistic of 0.44 (95% CI: 0.35, 0.53). This shows agreement similar to that in the emergency care sample. The Pearson Correlation Coefficient is 0.44 ($p<0.001$) and Cronbach's alpha is 0.61; while this alpha is below the acceptable cutoff of 0.7, as noted before, this measure is strongly influenced by the number of items being assessed, which may be driving this value.

Table 2. Two-way Contingency Table comparing results from the health literacy prediction model to the REALM-R

The table below compares health literacy as predicted by the health literacy prediction model that includes demographic factors and SILS to the REALM-R in the primary care sample

		Health Literacy as predicted by the Model	
		Inadequate Health Literacy	Adequate Health Literacy
Health Literacy as predicted by the REALM-R	Inadequate Health Literacy	107 (26.2%)	68 (16.7%)
	Adequate Health Literacy	42 (10.3%)	191 (46.8%)

CONCLUSION

A validated prediction model for inadequate health literacy that utilizes self-reported data would provide a feasible way to assess health literacy in fast-paced clinical settings. This would allow us to reach patients with limited health literacy with educational interventions and better meet their information needs. We conclude that the model created in the emergency department sample predicts limited health literacy, as assessed by the REALM-R, and validated this model in a sample obtained in a safety net primary care setting. SAS/STAT® provides useful tools for creating and validating prediction models. The process and code provided in this technical report has broad generalizability and can be adapted to develop and validate prediction models for other outcomes and in other settings.

REFERENCES

1. Health Literacy. (2014, January 9). Retrieved February 15, 2014, from http://www.cdc.gov/healthliteracy/

2. Gazmararian J a, Baker DW, Williams M V, et al. Health literacy among Medicare enrollees in a managed care organization. *JAMA*. 1999;281(6):545–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1002211110022111.

3. Parker RM, Baker DW, Williams M V, Nurss JR. The Test of Functional Health Literacy in Adults: A new instrument for measuring patients' literacy skills. *J Gen Intern Med*. 1995;10:537–541.

4. Williams M V, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease: a study of patients with hypertension and diabetes. *Arch Intern Med*. 1998;158(2):166. Available at: http://www.ncbi.nlm.nih.gov/pubmed/94485559448555.

5. Johnson K, Weiss BD. How long does it take to assess literacy skills in clinical practice? *J Am Board Fam Med*. 2008;21(3):211–4. doi:10.3122/jabfm.2008.03.070217.

6. Carpenter CR, Kaphingst KA, Goodman MS, Lin MJ, Melson AT, Griffey RT. Feasibility and Diagnostic Accuracy of Brief Health Literacy and Numeracy Screening Instruments in an Urban Emergency Department. Pines J, ed. *Acad Emerg Med*. 2014;21(2):137–146. doi:10.1111/acem.12315.

7. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008;23(5):561–6. doi:10.1007/s11606-008-0520-5.

8. Bass PF, Wilson JF, Griffith CH. A shortened instrument for literacy screening. *J Gen Intern Med*. 2003;18(12):1036–1038.

9. Davis TC, Long SW, Jackson RH, et al. Rapid Estimate of Adult Literacy in Medicine: A shortened screening instrument. *Fam Med*. 1993;25:391–395.

10. Jastak S, Wilkinson GS. *The wide range achievement test-revised: administration manual*. Jastak Assessment Systems; 1984.

11. Chew LD, Bradley K a, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med*. 2004;36(8):588–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1534342115343421.

12. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008;23(5):561–6. doi:10.1007/s11606-008-0520-5.

13. Baker DW, Williams M V, Parker RM, Gazmararian JA, Nurss JR. Development of a brief test to measure functional health literacy. *Patient Educ Couns*. 1999;38:33–42.

14. Davis TC, Crouch MA, Long SW, et al. Rapid Assessment of Literacy Level of Adult Primary Care Patients. *Fam Med*. 1991;23(6):433–435.

15. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. *BMC Fam Pract*. 2006;7:21. doi:10.1186/1471-2296-7-21.

16. Cohen J, A coefficient of agreement for nominal scales, Educational Psychol. Measurement. 1960; 37–46.

17. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011 Jun 27;2:53–5.

18. Nunnally J, Bernstein L. Psychometric theory. New York: McGraw-Hill Higher, INC; 1994.

19. Gennarelli R, Goodman MS. Measuring Internal Consistency of Community Engagement using the APLHA option of PROC CORR. Northeast SAS Users Group 2013. 2013; Available from: http://www.nesug.org/Proceedings/nesug13/128_Final_Paper.pdf

20. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiol Camb Mass. 2005 Jan;16(1):73–81.

21. Youden WJ. Index for rating diagnostic tests. Cancer. 1950 Jan;3(1):32–5.

ACKNOWLEDGMENTS

The project is supported by funding from the Barnes Jewish Hospital Foundation, Siteman Cancer Center and Washington University School of Medicine. Lucy D'Agostino McGowan is supported by the Program to Eliminate Cancer Disparities (National Institutes of Health, National Cancer Institute grant U54CA153460). The work of Dr. Goodman is supported by the Barnes-Jewish Hospital Foundation, Siteman Cancer Center, National Institutes of Health, National Cancer Institute grant U54CA153460, Washington University School of Medicine (WUSM) and WUSM Faculty Diversity Scholars Program. The work of Dr. Kaphingst was supported by funding from the Barnes Jewish Hospital Foundation, Siteman Cancer Center and Washington University School of Medicine.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Lucy D'Agostino McGowan
Organization: Washington University School of Medicine
Address: 600 S. Euclid Avenue Campus Box 8100
City, State ZIP: St. Louis, MO 63110
Work Phone: 314-286-2372
Email: dagostinol@wudosis.wustl.edu