Variation in Estimates of Limited Health Literacy by Assessment Instruments and Non-Response Bias

Joan M. Griffin, PhD^{1,2}, Melissa R. Partin, PhD^{1,2}, Siamak Noorbaloochi, PhD^{1,2}, Joseph P. Grill, MS¹, Somnath Saha, MD, MPH³, Annamay Snyder, MPH¹, Sean Nugent, BA¹, Alisha Baines Simon, MS¹, Ian Gralnek, MD, MSHS⁴, Dawn Provenziale, MD, MS⁵, and Michelle van Ryn, PhD, MPH⁶

¹Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center, Minneapolis, MN, USA; ²Department of General Internal Medicine, University of Minnesota School of Medicine, Minneapolis, MN, USA; ³Columbia Center for the Study of Chronic, Comorbid Mental and Physical Disorders, Portland VA Medical Center, Portland, OR, USA; ⁴Department of Gastroenterology, Rappaport Faculty of Medicine, Technion Israel Institute of Technology, GI Outcomes Unit, Haifa, Israel; ⁵Center for Health Services Research in Primary Care, Durham VA Medical Center, Durham, NC, USA; ⁶Department of Family Medicine and Community Health, University of Minnesota School of Medicine, Minneapolis, MN, USA.

OBJECTIVES: This paper compares estimates of poor health literacy using two widely used assessment tools and assesses the effect of non-response on these estimates.

STUDY DESIGN AND SETTING: A total of 4,868 veterans receiving care at four VA medical facilities between 2004 and 2005 were stratified by age and facility and randomly selected for recruitment. Interviewers collected demographic information and conducted assessments of health literacy (both REALM and S-TOFHLA) from 1,796 participants. Prevalence estimates for each assessment were computed. Non-respondents received a brief proxy questionnaire with demographic and self-report literacy questions to assess non-response bias. Available administrative data for non-participants were also used to assess non-response bias.

RESULTS: Among the 1,796 patients assessed using the S-TOFHLA, 8% had inadequate and 7% had marginal skills. For the REALM, 4% were categorized with 6th grade skills and 17% with 7–8th grade skills. Adjusting for non-response bias increased the S-TOFHLA prevalence estimates for inadequate and marginal skills to 9.3% and 11.8%, respectively, and the REALM estimates for ≤6th and 7–8th grade skills to 5.4% and 33.8%, respectively.

CONCLUSIONS: Estimates of poor health literacy varied by the assessment used, especially after adjusting for non-response bias. Researchers and clinicians should consider the possible limitations of each assessment when considering the most suitable tool for their purposes.

KEY WORDS: health literacy; veterans; prevalence; measurement; non-response bias; REALM; S-TOFHLA.

J Gen Intern Med 25(7):675–81

DOI: 10.1007/s11606-010-1304-2

© Society of General Internal Medicine 2010

INTRODUCTION

Health literacy (HL) is commonly defined as “the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions.”¹ Poor HL is now considered a risk factor for ill health and mortality.^{2–6} The Institute of Medicine (IOM), in “A Prescription to End Confusion,”⁵ recommends widespread assessment of HL in order to monitor and reduce the negative health effects of poor HL.

To date, the most commonly used HL assessments measure the individual capacity for either reading fluency or recognition of medical vocabulary. The Test of Functional Health Literacy (TOFHLA)⁷ and its short form, the S-TOFHLA,⁸ are common HL assessments used in research and clinical practice to assess reading fluency of health materials. The Rapid Estimate of Adult Literacy in Medicine (REALM),⁹ the most widely used assessment, tests for recognition of medical vocabulary. Few studies have compared prevalence estimates based on these measures,⁷ leaving researchers, clinicians and health educators with little data to evaluate which instrument best fits their needs. Moreover, with some notable exceptions,^{10,11} most HL studies using these assessments have relied on convenience sampling; examined clinically or geographically specific populations, often in locations or sites where evidence suggests rates of poor HL would be high¹²; and, in spite of the stigma attached to poor literacy,¹³ have not examined the effect of non-response on findings. In addition to inhibiting the generalizability of study conclusions, these methodological limitations may mask biases these assessments have, especially across sub-groups.

In this study of US veterans, we administered the two common HL assessments: the REALM⁹ and the S-TOFHLA⁸ in order to determine if estimates varied by measurement approach. We computed overall estimates of inadequate, marginal and adequate HL for each measure and, using administrative and survey data, evaluated whether non-response biased our findings.

METHODS

Face-to-face interviews were conducted with 1,796 US veterans who receive primary care services either at the Minneapolis,

Received November 16, 2009

Revised February 1, 2010

Accepted February 8, 2010

Published online March 12, 2010

West Los Angeles, Durham, or Portland VHA Medical Centers. The Institutional Review Board at each site approved the study protocol. The four study sites were deliberately chosen because they had a high volume of primary care patients and provided demographic and regional variation.

Study population. Eligible patients included those who were scheduled to have at least one primary care visit during the study's 12 month recruitment period and did not suffer from a severe cognitive disorder (e.g., dementia, schizophrenia), as determined by medical records review and cognitive impairment screening test administered prior to the interview. Since the HL assessments require individuals to read written material, blind patients or those with severely impaired vision were also excluded.

Studies have shown that age is one of the strongest correlates of HL.^{5,14} To ensure enough variability to detect differences in HL, eligible patients were stratified by age (<50, 50–75, >75) and facility. In total, 4,868 eligible patients were randomly selected for recruitment.

Recruitment. Invitations were mailed to randomly selected patients at each site. Ten days later, study recruiters telephoned each potential participant to invite them to participate. Up to six contact attempts were made at different times of day. Patients

willing to participate scheduled a 1-hour research appointment and received \$25 at the interview for participating.

Study sample. Of the 4,868 veterans selected for recruitment, we were unable to contact 21% (Fig. 1). Of the 3,850 we were able to contact, 23% refused to participate because of scheduling or transportation difficulties, 9% for other reasons, and 4% were ineligible. Of those contacted, 64% agreed to an interview. Of those, 9% did not show up for their appointment or could not be rescheduled after missing their interview; 8% showed up, but did not complete the interview due to ineligibility or participant decision not to continue. We completed interviews with 53% (n = 1,796) of the eligible participants.

Non-response sample. To assess non-response bias, demographic characteristics (age, marital status, and self-reported race/ethnicity) and history of comorbid conditions were extracted from administrative and medical record data for all 4,384 randomly selected for recruitment and not found to be ineligible. Non-respondents included all eligible patients who could not be reached (n = 1,018), did not attend their scheduled research appointment (n = 325), initiated but did not finish the interview (n = 22), "soft refusers" (e.g., those who reported scheduling or transportation difficulties or were too busy;

Figure 1. Study participant flow diagram.

n=1,223), and “hard refusers” (e.g., those not interested in the research). All non-responders except for hard refusers were also mailed a proxy survey packet, described below.

Data Collection

Face-to-face interviews. Prior to administering each survey, interviewers screened patients for cognitive and visual impairments to verify eligibility. Cognitive impairment was assessed using the Mini-Cog, a validated brief screening test for dementia.¹⁵ Those testing positive for dementia (Mini-Cog score <3) were ineligible. Visual acuity was assessed using standard vision charts. Patients with corrected visual acuity of 20/100 or worse were ineligible because poor eyesight might confound the HL assessments. After visual and cognitive screening, interviewers administered the study’s survey to all eligible participants.

Non-response proxy survey. Surveys were sent after all face-to-face interviews were completed. The survey packet included a small cash incentive, cover letter, self-addressed stamped envelope and a ten-item questionnaire containing four self-report HL questions (described below) and six demographic questions (marital status, race, ethnicity, education, employment and income).¹⁶ Reminder postcards were mailed 1 week after the first mailing. A second questionnaire was mailed to anyone who did not return a blank or completed survey within 3–4 weeks of the first mailing. A total of 1,435 (64%) of the 2,237 participants who did not complete a face-to-face interview completed this proxy questionnaire.

Measures

Health literacy assessments. The REALM and S-TOFHLA were used to assess HL skills. The REALM measures HL by assessing the correct pronunciation of 66 common medical terms. It is strongly correlated with other standardized reading assessments and has excellent intra-subject reliability.^{9,17,18} Scores are most often categorized into grade levels ($\leq$ 3rd grade, 4–6th grade, 7–8th grade, high school). Given the distribution of scores and their correlation to S-TOFHLA scores, we recoded REALM scores into these same categories: $\leq$ 6th grade (0–44 words pronounced correctly), 7–8th grade (45–60) and high school (61–66).

The S-TOFHLA includes 4 numeracy items followed by 36 reading comprehension items. The reading comprehension passages have good internal consistency.^{7,8} To assess numeracy, participants were given props and asked questions that require them to make numerical calculations and interpret meanings from test results and recognize an appointment time. The reading comprehension assessment includes two reading passages of varying difficulty, with every third to fourth word of the passage removed and replaced with a list of four possible words from which to choose to complete the sentence. Participants are timed and given a total of 7 min (the standard protocol for the S-TOFHLA)¹⁹ to complete the assessment. Total scores range from 0–100 and are then categorized into inadequate, marginal and adequate HL skills using established cutoff scores (inadequate, 0–53; marginal, 54–66; adequate, 67–100),¹⁹ making categories comparable to those used for REALM. During the interview, the REALM was administered first, followed by the numeracy and then the timed reading comprehension sections of the S-TOFHLA.

Four self-report HL questions, shown to be adequately sensitive and specific in predicting HL by level of REALM and S-TOFHLA,

were also administered during the interviews, preceding the administration of the REALM and S-TOFHLA. These same questions were also included in the proxy survey. Three questions were adapted from items developed by Chew.²⁰ We added another question: “How often do you have problems learning about your medical condition?” Responses to the self-report HL questions were scored on a 5-point Likert scale. These questions have been shown to be adequately sensitive and specific in predicting HL by level of REALM and S-TOFHLA.²¹

Covariates. Prior studies have described a common set of factors associated with both inadequate literacy and HL skills, most of which are tied to poverty status.^{5,22} Therefore, demographic, socioeconomic and health characteristics were collected during the interview and from administrative data files. Because a high burden of morbidity could confound estimates of adequate HL, chronic disease and mental health history for each patient was also extracted from medical records. Chronic diseases were summarized using a Charlson comorbidity score.²³ Mental health diagnoses were categorized into three groups: (1) no mental health diagnoses, (2) single psychiatric or substance abuse related diagnosis, or (3) dual diagnosis (psychiatric and substance abuse).

Statistical Analyses

Constructing estimates of health literacy. To account for our complex sampling design, we used stratified weighted estimation methods to estimate HL levels.²⁴ For each assessment, stratified estimates were calculated using all participants who completed the test during interviews.

Adjusting for non-response bias. Multiple imputation (MI) was used to adjust for non-response bias. The MI procedure in SAS 9.1 and a logistic regression model were used to compute and assign a probability to each of the possible values that could be assumed by a variable whose value is missing for any given case. The probabilities assigned to the replacement values depended on the values assumed by all other covariates in the model for the case.²⁵ A number randomly chosen between 0 and 1 from a uniform probability distribution was compared to the assigned probabilities to select the imputed value. Covariates in the model included administrative data (e.g., age and gender) and self-report HL items. Imputation was done iteratively to maintain a monotone missing pattern. The initial covariates with no missing values were used to impute replacement values for the item with some but fewer missing values than any of the remaining items. This item was then added to the set of covariates in the logistic model, which was used to impute values for the item with the second fewest number of missing values, and so on. The missing S-TOFHLA and REALM scores for non-respondents were imputed last using a linear regression model with all the initial and imputed items as covariate predictors. Once the missing S-TOFHLA and REALM scores had been replaced, they were classified as inadequate, marginal or adequate.

This imputation process was repeated five times, creating five different but complete data sets. The prevalence of each HL level was computed for each of the five data sets, and these results were combined to create one overall set of non-response-adjusted HL prevalence estimates, accounting for data variability due to imputation and sampling.²⁶

RESULTS

Sample characteristics. Characteristics of the study sample are described in Table 1. Compared to interview participants, proxy survey responders and non-responders had lower levels of education and household income. Non-responders were also more likely to be younger, never married, African American, live in urban areas and have a mental health diagnosis than interview participants.

Prevalence of HL. The variation across groups in Table 1 suggests our estimates might be influenced by response bias; therefore, unadjusted and adjusted estimates (corrected for response bias) are shown in Table 2. Because we had survey data for proxy survey responders, but not non-responders, proxy survey responders had more observed data to use for imputation and substantially fewer missing values for variable items typically associated with adequate HL. Proxy survey responders, therefore, had relatively more accurate imputed S-TOFHLA and REALM scores, and more accurate adjusted prevalence estimates.

Among the 1,789 respondents who completed the REALM, 3.9% scored at or below the 6th grade reading level and 17.3% scored at 7th to 8th grade levels (Table 2). Using just proxy survey responders to adjust for bias, estimates increased to 5.2% for ≤ 6 th grade and 30.5% for 7th to 8th grade. After adjusting for all non-respondents (proxy and non-responders), estimates increased to 5.4% and 33.8%, respectively. Among patients who completed the S-TOFHLA, 8.1% were classified as having inadequate HL, and 7.4% had marginal skills (Table 2). These estimates increased when proxy survey data were used to adjust for bias: 9.1% were classified as having inadequate skills, and 10.7% had marginal skills. Further adjustment using all non-responders further increased estimates, to 9.3% and 11.8%, respectively.

DISCUSSION

Both the IOM's report⁵ and Healthy People 2010²⁷ conclude that improving HL is a national priority and may be critical for reducing disease and health disparities. Having valid and reliable HL assessments is vital for understanding and reducing the negative effects of poor HL. The objective of this study was to compare prevalence estimates derived from the two most common HL assessments and determine the effect of non-response on those estimates. While other measures of HL exist,^{20–22,28} the REALM and S-TOFHLA are the most commonly used in clinical and community settings, yet few studies have compared them^{7,8}. In the literature, however, they are often considered equivalent. A recent systematic review¹², for example, reports similar prevalence rates between the set of studies using either the TOFHLA or S-TOFHLA and those using REALM. Our study, which directly compares the two, brings this into question.

We draw three important conclusions from our results. First, the prevalence of low HL varies by the assessment used. This finding has important implications for health systems and researchers as they take up the IOM recommendations⁵ to conduct HL assessments locally and nationally in order to determine the magnitude of poor HL, monitor how it changes over time and find innovative ways to improve it. In the two prior studies directly comparing the REALM and S-TOFHLA,^{7,8}

estimates were strongly correlated ($r=0.80$), and agreement was strongest for those with the highest and lowest skills, but differed significantly in the middle ranges of the tests. In our study, nearly two times as many were categorized with inadequate skills using S-TOFHLA than with REALM and three times as many were categorized with marginal (7–8th grade) by REALM than with S-TOFHLA. Differences across assessments could indicate that one assessment is less accurate than the other, especially for certain thresholds, or that parameters are less stable across different demographic groups. Another explanation, however, is that each instrument measures different components of individual capacity for understanding health-related information and, thus, they are not comparable instruments. Baker has suggested that print literacy is related to two constructs: reading fluency (prose, quantitative and document fluency) and prior knowledge (vocabulary and conceptual knowledge of health and health care).²⁹ It is possible that the S-TOFHLA measures reading fluency more accurately, whereas the REALM measures prior knowledge more accurately. Understanding the conceptual differences in these assessments may be helpful to researchers and practitioners who are trying to determine which assessment(s) is most appropriate for measuring an intervention's progress. Because the correlation between the two assessments is strong, any program designed to improve health literacy as measured by one assessment tool would likely show benefits in the other. However, existing programs may want to use both assessments in the development and evaluation of HL interventions in order to assure that both reading fluency and knowledge skills are developed, thereby maximizing the impact on overall HL skills.

Our second conclusion is that non-response bias affects prevalence estimates and that estimates based on REALM are more affected by response bias than S-TOFHLA. Most HL studies to date have relied on convenience samples, which can compromise the validity of findings by inflating estimates of poor HL. Random sampling can help to improve the accuracy of estimates in large populations, but because disenfranchised or stigmatized groups are often less likely to participate in research^{13,16,30–33}, these estimates may also be biased. Two prior HL studies, one a large study of Medicare enrollees in a national managed care organization that used the S-TOFHLA¹¹ and the other, the National Assessment of Adult Literacy (NAAL), which used an instrument developed specifically for that study,²² have assessed non-response bias. Both used census-tract data and found that nonresponders were not more likely to come from disenfranchised areas or areas with high concentrations of people at risk for poor HL, but instead from areas with higher incomes and educational attainment, and less concentration of blacks or minority race. While the NAAL data were weighted to adjust for non-response bias, the Medicare enrollee study was not, and thus, estimates for poor HL may have been deflated due to bias. These results and our own findings suggest that non-response bias is an important consideration for all HL studies. Well-designed qualitative studies may be particularly helpful for researchers to assess the validity of the assumptions used in our non-response analyses, and further quantitative studies using randomized samples of patients are needed to understand how bias differs by assessment and demographic sub-groups.

Finally, the prevalence of the poorest HL skills in this study is lower than in other large studies, although a notable proportion of participants (4–9%) still have only the most

Table 1. Demographic and Health Characteristics of Study Participants by Response Status

Response status of participants								
Characteristics	Responding to face-to-face interview (n=1,796)		Not responding to face-to-face survey but responding to proxy ¹ survey (n=1,435)		Not responding to either face-to-face or proxy survey (n=1,153) ^a		χ ^{2a}	p≤
	N		N	%	N	%		
Demographic								
Gender								
Female	246	13.7	194	13.5	195	16.9	7.47	0.02
Male	1,550	86.3	1241	86.5	958	83.1		
Missing	0		0		0			
Age								
<50	537	29.9	418	29.1	464	40.2	87.82	<0.01
50–75	748	41.7	490	34.2	317	27.5		
>75	511	28.4	527	36.7	372	32.3		
Missing	0		0		0			
Education								
Grade 11 or less	164	9.1	210	14.6	169	14.7	43.61	<0.01
High school grad	447	24.9	374	26.1	343	29.7		
Some college/tech deg	727	40.5	458	31.9	434	37.6	70.32	<0.01
College grad	288	16.0	214	14.9	145	12.6		
Post grad (MS, PhD)	159	8.9	87	6.1	62	5.4		
Missing/unknown	11	0.6	92	6.4	0			
Income								
Under \$20,000	613	34.1	777	54.2	799	69.3	425.40	<0.01
\$20,000–\$40,000	647	36.0	418	29.1	258	22.4		
\$40,000–\$60,000	280	15.6	136	9.5	41	3.5		
Over \$60,000	256	14.3	104	7.2	39	3.4		
Missing	0		0		16	1.4		
Marital status								
Married/live together	1,048	58.3	777	54.2	426	37.0	173.43	<0.01
Divorced/separated	402	22.4	324	22.6	336	29.1		
Widowed	138	7.7	142	9.9	91	7.9		
Never married	208	11.6	191	13.3	286	24.8		
Missing	0		0		14	1.2		
Race								
White	1,299	72.3	966	67.3	492	42.7	174.15	<0.01
Black	334	18.6	325	22.6	362	31.4		
Other	156	8.7	140	9.8	33	2.9		
Missing	7.4		4	.3	266	23.0		
Urban status								
Urban	1,142	63.6	981	68.4	809	70.1	37.48	<0.01
Suburban/exurban	535	29.8	374	26.0	319	27.7		
Rural	119	6.6	80	5.6	25	2.2		
Missing	0		0		0			
Site								
SE	418	23.3	265	18.5	357	31.0	240.83	<0.01
Midwest	488	27.2	354	24.7	127	11.0		
NW	431	24.0	262	18.2	168	14.6		
W	459	25.5	554	38.6	501	43.4		
Missing	0		0		0			
Health								
Physical comorbidities								
0	783	43.6	600	41.8	478	41.5	9.64	0.14
1	372	20.7	301	21.0	254	22.0		
2	291	16.2	233	16.2	156	13.5		
3 or more	350	19.5	301	21.0	265	23.0		
Missing	0		0		0			
Mental health (MH)							21.99	<0.01
No MH diagnosis	831	46.3	734	51.1	499	43.3		
Psych diagnosis	166	9.2	118	8.2	95	8.2		
Substance use diagnosis	519	28.9	387	27.0	345	29.9		
Dual diagnosis	280	15.6	196	13.7	214	18.6		
Missing	0		0		0			

Bolded frequencies are based exclusively on imputed values

^aBolded chi-square statistic and corresponding p-value summarize the association between those responding to the face-to-face interviews and those not responding to the face-to-face interview or proxy survey. All other chi-square statistics and p-values correspond to tests of the null hypothesis that the distribution of levels for each demographic or health characteristic is the same for all three groups of participants (face-to-face vs. proxy respondents vs. non-respondents)

Table 2. Stratified Point and Interval (95% CI) Estimates of the Prevalence (%) of HL Categories Based on REALM and S-TOFHLA

Analysis	n	REALM (reading grade level) ^a		
		≤6th grade (score=0–44)	7th–8th grade (score=45–60)	High school (score=61–66)
Group 1 only	1,796	3.86 (2.83, 4.90)	17.27 (15.21, 19.34)	78.47 (76.24, 80.69)
Non-response adjustment I ^b	3,231	5.16 (3.71, 6.60)	30.46 (28.04, 32.88)	64.38 (62.07, 66.70)
Non-response adjustment II ^c	4,384	5.39 (3.75, 7.03)	33.84 (31.06, 36.62)	60.77 (58.07, 63.47)
		S-TOFHLA (literacy category)		
		Inadequate (score=0–53)	Marginal (score=54–66)	Adequate (score=67–100)
Group 1 only	1,796	8.12 (6.59, 9.64)	7.39 (5.92, 8.86)	84.50 (82.51, 86.48)
Non-response adjustment I ^b	3,231	9.11 (7.48, 10.73)	10.73 (9.31, 12.14)	80.17 (78.29, 82.04)
Non-response adjustment II ^c	4,384	9.26 (7.73, 10.79)	11.82 (6.99, 9.85)	78.92 (76.67, 81.17)

^aScores for the REALM test were not recorded for 7 of the 1,796 responding participants. Their REALM scores were imputed along with the REALM scores of those in proxy respondents group because they provided information via the face-to-face interview that exactly matches the information provided by those in the proxy respondent group

^bEstimates using face-to-face respondents and proxy respondents = 1,435

^cEstimates using face-to-face, proxy and nonrespondents

rudimentary skills. Our estimates of poor HL differ from a convenience sample of other primary care veteran patients, where 10–15% had skills at ≤6th grade³⁴, but similar to other veteran samples, including those undergoing cancer screening (where 36% had skills at ≤8th grade)³⁵ and those in a preoperative clinic (4.5% had inadequate; 7.5% had marginal skills).²⁰ Our estimates of poor HL were also lower than those in studies of non-veterans. In the NAAL, the most comprehensive, nationally representative study of English-language literacy and HL in the US, 14% had below basic skills,²² and in a recent systematic review of HL studies, 26% had inadequate skills.¹² A number of factors could explain why our estimates of poor HL are lower than those from prior studies. First, veterans using VHA may have different health care experiences, including patient education, than non-veterans. Second, variation may reflect the differences across studies, including the assessment tools used. The NAAL, for example, did not use the REALM or S-TOFHLA, making comparisons difficult. Third, study methodology may lead to difference in estimates. Many prior studies based on smaller, single site convenience samples. Fourth, population characteristics may differ across studies. Veteran's level of English fluency and cognitive capacity to function in English, for example, may be higher than non-veterans. Although we did not assess whether English was their primary language as part of this study, veterans are required to demonstrate the ability to speak and write English at a functional level as a requirement for entering the armed forces, and are screened for physical or mental conditions that might inhibit their ability to serve.^{36,37} Education is also strongly linked to literacy^{22,38}, and veterans in this sample have relatively high levels of education compared to other studies. In the systematic review of HL studies, for example, the majority of studies reported 35–55% of their samples having less than a high school education, while we found just under 10%.¹² Variation in prevalence estimates, therefore, may be due to both methodological and population differences. Future research efforts, especially analyses of the NAAL data that include veteran status (but not specifically veterans using VHA), could be used to further assess demographic differences in HL and non-response bias.

Our findings are tempered by a number of limitations. Accepted definitions of HL emphasize the dynamic communication and negotiation strategy between an individual and the health care environment. Our assessments, however, were conducted privately, free of interruptions or distractions, using assessments with a narrower focus than the broader, accepted definition. These estimates, therefore, may underestimate how well people function in ordinary health encounters, since encounters often include complex information in different formats (i.e., written, oral, non-verbal) and distractions that can inhibit comprehension and recall. Related to this, our overall response rate was low compared to other studies that used convenience sampling, but relatively consistent with studies that used random sampling. Possible bias due to non-response, however, was addressed. It is also important to note that our study and the literature cited were conducted in the US and our findings may not generalize to other English-speaking countries where different factors may affect HL and non-response. Finally, in spite of state-of-the-art imputation methods, imputation is never perfect, and therefore, estimates adjusted for bias may actually underestimate the true prevalence of poor HL.

CONCLUSIONS

Estimates of poor health literacy based on the REALM and S-TOFHLA differ, especially among those with the equivalent of marginal skills and after adjusting for non-response bias. Researchers and clinicians who direct health systems efforts to improve HL are advised to consider the possible limitations of each assessment when considering the most suitable tool for their purposes.

Acknowledgements: This research was supported by the Department of Veterans Affairs, including a grant from VA Health Services Research and Development Service (CRI-03-151-1). Dr. Griffin also received support as a VA Merit Review Entry Program (MREP) awardee. Dr. Gralnek was supported by a VA HSR&D Advanced Research Career Development Award and IIS 01-191-1. The views expressed in this article are those of the author(s) and do not

necessarily represent the views of the Department of Veterans Affairs or the United States Government. Part of this work was presented at the 24th Annual VA Health Services Research Meeting, Crystal City, VA, February 17, 2006. The authors wish to thank all the veteran participants for their time and the study interviewers for their commitment to this project.

Conflict of Interest: None disclosed.

Corresponding Author: Joan M. Griffin, PhD; Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center (152/3E-109), One Veterans Drive, Minneapolis, MN 55417, USA (e-mail: joan.griffin2@va.gov).

REFERENCES

1. **Ratzan SC.** Health literacy: Communication for the public good. *Health promot Int.* 2001;16:207–14.
2. **Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al.** Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health.* 2002;92:1278–83.
3. **Baker DW, Parker RM, Williams MV, Clark WS, Nurss J.** The relationship of patient reading ability to self-reported health and use of health services. *A J Public Health.* 1997;87:1027–30.
4. **Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J.** Health literacy and mortality among elderly persons. *Am J Public Health.* 2007;167(14):1503–9.
5. **Nielsen-Bohlman LT, Panzer AM, Kindig DA.** Health Literacy: A prescription to end confusion. Washington, DC: Institute of Medicine of the National Academies, The National Academies Press; 2004.
6. **DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP.** Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med.* 2004;19:1228–39.
7. **Parker RM, Baker DW, Williams MV, Nurss JR.** The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med.* 1995;10:537–41.
8. **Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss JR.** Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999;38:33–42.
9. **Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, et al.** Rapid assessment of literacy levels of adult primary care patients. *Fam Med.* 1991;23:433–35.
10. **Bennett IM, Kripalani S, Weiss BD, Coyne CA.** Combining cancer control information with adult literacy education: opportunities to reach adults with limited literacy skills. *Cancer Control.* 2003;10(5 Suppl):81–3.
11. **Gazmararian JA, Baker DW, Williams MV, Parker RM, Scott TL, Green DC, et al.** Health literacy among Medicare enrollees in a managed care organization. *JAMA.* 1999;281:545–51.
12. **Paasche-Orlow M, Parker R, Gazmararian J, Nielsen-Bohlman L, Rudd R.** The prevalence of limited health literacy. *J Gen Intern Med.* 2005;20:175–84.
13. **Pariikh NS, Parker RM, Nurss JR, Baker DW, Williams MV.** Shame and health literacy: The unspoken connection. *Patient Educ Couns.* 1996;27:33–39.
14. **Baker DW, Gazmararian JA, Sudano J, Patterson M.** The association between age and health literacy among elderly persons. *J Geront, Ser. B Psychol Sci Soc Sci.* 2000;55:S368–74.
15. **Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A.** The Mini-Cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry.* 2000;15:1021–27.
16. **Sheikh K, Mattingly S.** Investigating non-response bias in mail surveys. *J Epidemiol Community health.* 1981;35(4):293–6.
17. **Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al.** Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med.* 1993;25:391–95.
18. **Davis TC, Michielutte R, Askov EN, Williams MV, Weiss BD.** Practical assessment of adult literacy in health care. *Health Educ Behav.* 1998;25:613–24.
19. **Nurss J, Parker R, Williams MV, Baker DW.** TOFHLA-Test of Functional Health Literacy in Adults Manual. Camp, NC: Peppercorn Books and Press; 2001.
20. **Chew LD, Bradley KA, Boyko EJ.** Brief questions to identify patients with inadequate health literacy. *Fam Med.* 2004;36:588–94.
21. **Chew LD, Griffin JM, Partin M, Noorbaloochi S, Grill J, Snyder A, et al.** Validation of screening questions for limited health literacy in a large outpatient population. *J Gen Intern Med.* 2008;23:561–66.
22. **Kutner M, Greenberg E, Jin Y, Paulsen C.** The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483). Washington, DC: National Center for Education Statistics, US Department of Education; 2006.
23. **Deyo R, Cherkin D, Ciol M.** Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45:613–19.
24. **Cochran W.** Sampling Techniques. 3rd ed. New York: Wiley & Sons; 1977.
25. **Allison PD.** Missing Data. Thousand Oaks, CA: Sage Publications; 2001.
26. **Rubin D.** Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons, Inc.; 1987.
27. **US Department of Health and Human Services.** Healthy People 2010. Understanding and Improving Health. 2nd ed. Washington, DC: US Government Printing Office; 2000.
28. **Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al.** Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med.* 2005;3(6):514–22.
29. **Baker DW.** The meaning and the measure of health literacy. *J Gen Intern Med.* 2006;21:878–83.
30. **Sheikh K.** Predicting risk among non-respondents in prospective studies. *Eur J Epidemiol.* 1986;2(1):39–43.
31. **Cottler LB, Zipp JF, Robins LN, Spitznagel EL.** Difficult-to-recruit respondents and their effect on prevalence estimates in an epidemiologic survey. *A J Epidemiol.* 1987;125(2):329–39.
32. **McNutt LA, Lee R.** Intimate partner violence prevalence estimation using telephone surveys: understanding the effect of nonresponse bias. *A J Epidemiol.* 2000;152(5):438–41.
33. **Hardy RE, Ahmed NU, Hargreaves MK, Semanya KA, Wu L, Belay Y, et al.** Difficulty in reaching low-income women for screening mammography. *J Health Care Poor Underserved.* 2000;11(1):45–57.
34. **Kelly PA, Haidet P.** Physician overestimation of patient literacy: A potential source of health care disparities. *Patient Educ Couns.* 2007;66:119–22.
35. **Dolan NC, Ferreira MR, Davis TC, Fitzgibbon ML, Rademaker A, Liu D, et al.** Colorectal cancer screening knowledge, attitudes, and beliefs among veterans: does literacy make a difference? *J Clin Oncol.* 2004;22:2617–22.
36. **Eitelberg MJ, Laurence JH, Waters BK, Perelman LS.** Screening for Service: Aptitude and Education Criteria for Military Entry. Office of the Naval Research. HumRRO FR-PRD-83-24. Alexandria, VA: Human Resources Research Organization; 1984.
37. **Sticht, TG and Armstrong, WB.** *Adult Literacy in the United States.* 1994. Retrieved February 8, 2010, from <http://www.nald.ca/library/research/adlitUS/cover.htm>.
38. **Kirsch I, Jungeblut A, Jenkins L, Kolstad A.** Adult literacy in America: A first look at the results of the National Adult Literacy Survey. Washington, DC: National Center for Education, US Department of Education; 1993.