

## REVIEWS

# Different Measures, Different Outcomes? A Systematic Review of Performance-Based versus Self-Reported Measures of Health Literacy and Numeracy

Eric S. Kiechle, MD MPH<sup>1</sup>, Stacy Cooper Bailey, PhD MPH<sup>2</sup>, Laurie A. Hedlund, MA<sup>3</sup>, Anthony J. Viera, MD MPH<sup>1,4,5</sup>, and Stacey L. Sheridan, MD MPH<sup>1,5,6,7</sup>

<sup>1</sup>Program on Health Care and Prevention, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA; <sup>2</sup>Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA; <sup>3</sup>Health Literacy and Learning Program, Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; <sup>4</sup>Department of Family Medicine, University of North Carolina, Chapel Hill, NC, USA; <sup>5</sup>Cecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, NC, USA; <sup>6</sup>Division of General Medicine and Clinical Epidemiology, Department of Medicine, University of North Carolina, Chapel Hill, NC, USA; <sup>7</sup>University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

**BACKGROUND:** Health literacy (HL) and numeracy are measured by one of two methods: performance on objective tests or self-report of one's skills. Whether results from these methods differ in their relationship to health outcomes or use of health services is unknown.

**METHODS:** We performed a systematic review to identify and evaluate articles that measured both performance-based and self-reported HL or numeracy and examined their relationship to health outcomes or health service use. To identify studies, we started with an AHRQ-funded systematic review of HL and health outcomes. We then looked for newer studies by searching MEDLINE from 1 February 2010 to 9 December 2014. We included English language studies meeting pre-specified criteria. Two reviewers independently assessed abstracts and studies for inclusion and graded study quality. One reviewer abstracted information from included studies while a second checked content for accuracy.

**RESULTS:** We identified four "fair" quality studies that met inclusion criteria for our review. Two studies measuring HL found no differences between performance-based and self-reported HL for association with self-reported outcomes (including diabetes, stroke, hypertension) or a physician-completed rheumatoid arthritis disease activity score. However, HL measures were differentially related to a patient-completed health assessment questionnaire and to a patient's ability to interpret their prescription medication name and dose from a medication bottle. Only one study measured numeracy and found no difference between performance-based and self-reported measures of numeracy and colorectal cancer (CRC) screening utilization. However, in a moderator

analysis from the same study, performance-based and self-reported numeracy were differentially related to CRC screening utilization when stratified by certain patient-provider communication behaviors (e.g., the chance to always ask questions and get the support that is needed).

**DISCUSSION:** Most studies found no difference in the relationship between results of performance-based and self-reported measures and outcomes. However, we identified few studies using multiple instruments and/or objective outcomes.

**KEY WORDS:** health literacy; literacy; measures; measurement; numeracy.

J Gen Intern Med 30(10):1538–46

DOI: 10.1007/s11606-015-3288-4

© Society of General Internal Medicine 2015

## INTRODUCTION

Low levels of health literacy and numeracy have been associated with a number of negative health outcomes, including higher mortality in seniors, increased use of emergency departments and inpatient facilities, and lower use of some preventive services.<sup>1</sup> However, no gold standard currently exists for measuring health literacy. Researchers have raised concerns that measures from existing instruments may be measuring different underlying constructs.<sup>2</sup> Further, experts have recommended using multiple measures of health literacy to learn more about how measures perform against each other and to better quantify the relationships between measures and health outcomes.<sup>3</sup>

The instruments most often used to measure health literacy and numeracy in clinical studies are the Short Test of Functional Health Literacy in Adults (S-TOFHLA), the Rapid Estimate of Adult Literacy in Medicine (REALM), and the Schwartz and Woloshin numeracy questions.<sup>4,5</sup> All are performance-based, or objective, in their assessments. The S-TOFHLA requires patients to select one of four words to fit into 36 blanks scattered through two medical passages and

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11606-015-3288-4) contains supplementary material, which is available to authorized users.

---

Received March 3, 2014

Revised July 31, 2014

Accepted March 10, 2015

Published online April 28, 2015

complete a four tasks testing numerical ability, while the REALM assesses pronunciation of 66 medical words of varying difficulty. The Schwartz and Woloshin questions,<sup>6</sup> and the similar Lipkus numeracy scale,<sup>7</sup> ask participants to perform such tasks as predicting the behavior of a perfect coin, converting probabilities into percentages, and vice versa.

Self-reported health literacy and numeracy instruments (i.e., those that ask patients to self-rate their abilities) are increasingly common.<sup>8–14</sup> Frequently used self-report measures include Chew et al.'s brief validated screening questions (BSQ), the Single Item Literacy Screener, and the Subjective Numeracy Scale.<sup>8,9,13</sup> All involve patients describing themselves and their preferences or skills. Most of these subjective measures have been designed for screening rather than measuring health literacy

or numeracy in clinical settings, and have been validated against objective instruments (Table 1). They also have the advantage of being shorter and potentially less embarrassing for patients.<sup>8,18,19</sup> They therefore could allow for more efficient research about health literacy, as well as fewer negative feelings for patients involved in health literacy research. However, it is currently not clear whether self-reported measures have the same relationship to outcomes as the performance-based measures. Self-reported and performance-based measures differ in many potentially important ways (e.g., their intent, length, psychometric properties) that could affect their relationship with outcomes.<sup>20,21</sup>

No reviews to date have examined whether performance-based and self-reported measures of health literacy and numeracy have the same relationship to health outcomes when

**Table 1. Summary of Existing Self-Report Health Literacy and Numeracy Measures**

Instrument	Description	Scale	Categorization	Performance
Brief Screening Questions (BSQ) <sup>8</sup>	Three questions (confidence with medical forms, problems with reading/ understanding health information, need for help with these materials) to screen for limited health literacy	Each question scored on a 5-point Likert scale. Scale wording varies based on question	Validating study does not suggest a cutpoint, stating instead that cutoffs should depend on purpose and prevalence of limited HL in a given environment	Performance assessed against S-TOFHLA with area under ROC curve (AUROC) values of 0.76 to 0.87, depending on the question
Single Item Literacy Screen (SILS) <sup>9</sup>	Single question developed from three questions (BSQ) identified to be effective at screening for inadequate health literacy. Questions ask about frequency with which patients need help reading medical materials	5-point Likert scale, from Never to Always	Responses of "Sometimes" or more considered positive, classifying patient as at-risk for difficulty with reading printed health materials	Performance assessed against S-TOFHLA with AUROC of 0.73 (95 % CI 0.69–0.78) for inadequate or marginal HL
Brief Health Literacy Screening Tool (BRIEF) <sup>12</sup>	Addition of oral HL question (difficulty understanding what is told to patient) to three BSQs to create a four-question scale	Each question scored on 5-point Likert scale, summed for score range of 4–20. Scale wording varies by question.	Scores of 17–20 considered adequate HL, scores of 13–16 considered marginal HL and scores < 13 considered inadequate HL	Performance assessed against S-TOFHLA with AUROC curve 0.74 (95 % CI 0.67 to 0.80) and against REALM with AUROC 0.69 (95 % CI 0.64–0.75)
eHealth Literacy Scale (eHEALS) <sup>10</sup>	Eight-item measure of electronic health literacy designed to assess patients' skills at using information technology for health purposes	Each question is on a 5-point Likert scale, from Strongly Disagree to Strongly Agree	No categorization stated	Performance not evaluated against other measures
All Aspects of Health Literacy Scale (AAHLS) <sup>14</sup>	Fourteen questions encompassing all three domains of HL (functional, communicative and critical*) reflecting skills needed in these domains	Questions are on a 3-point scale (Never, Sometimes, Often), except for two dichotomous questions	No categorization stated	Performance not evaluated against other measures
Health Literacy Management Scale (HeLMS) <sup>15</sup>	Twenty-nine-item scale divided into eight domains measuring individuals' abilities and broader social/ environmental contexts	Each item is on a 5-point Likert scale, from 'Unable to do' to 'Without any Difficulty'	Reported as mean score for each domain, no cutoff provided in validation study	Performance not evaluated against other measures
Subjective Numeracy Scale (SNS) <sup>13</sup>	Eight-item measure with questions regarding beliefs about skill in mathematical operations and about preference of presentation of numerical information	Each question is on a 6-point Likert scale. Scale wording varies by question. SNS score is calculated from the mean of individual scores, range of 1–6	No categorization stated	Performance assessed against Lipkus numeracy scale, with adjusted correlation coefficient $r=0.68$ ( $p<0.01$ )
STAT Confidence Scale <sup>16</sup>	Three questions assessing problems understanding and confidence interpreting medical statistics	Answers on Likert scales (Very easy to Very hard, and Strongly disagree to Strongly agree) converted to a 100-point scale	No categorization stated	Only weakly correlated with a performance-based medical data interpretation test developed by the authors ( $r=0.15$ , $p=0.04$ )

\*Functional literacy describes basic skills in reading and writing to function in everyday situations. Communicative literacy includes more advanced cognitive and literacy skills that combine with social skills to enable participation. Critical literacy encompasses advanced cognitive skills needed to critically analyze information and use the information to exert greater control over situations.<sup>17</sup>

measures are applied within the same samples. An understanding of this issue is important because differential relationships could explain discrepant findings in systematic reviews and individual research studies. To explore this issue, we examined quantitative studies that compared performance-based with self-reported health literacy or numeracy across a range of health outcomes.

## METHODS

### Data Sources and Selection

We started our review by examining studies included in a 2011 systematic evidence review funded by the Agency for Health Research and Quality (AHRQ).<sup>22</sup> This review is the most comprehensive assessment of the relationship between health literacy and numeracy and health outcomes to date and considered the relationship of both print literacy and numeracy with health outcomes, including knowledge (only for numeracy), accuracy of risk perception, skills, use of health services, disease severity, quality of life, mortality, and costs. We continued our review by searching MEDLINE using the same search string as the AHRQ review (Table 2) to identify newer studies that examined the relationship of both performance-based and subjective measures of health literacy and health outcomes. We did not search other databases, given their low yield in prior work (< 7 % of all articles identified in the AHRQ review were

outside of MEDLINE).<sup>23</sup> The start date for our search was 1 year prior to the search end date used in the AHRQ review to capture articles that may not yet have been indexed at the time of that review (1 February 2010); we updated our search through 9 December 2014. We also hand-searched reference lists of included studies and of a systematic review of available health literacy measures for additional studies.<sup>2</sup>

Inclusion and exclusion criteria were modeled on the 2011 systematic review of HL and health outcomes by Berkman et al.<sup>22</sup> (Table 3). We included English language studies of any observational or experimental study design and excluded qualitative studies, validation studies, narrative review articles, case reports, editorials and letters. We also newly required that studies had to measure health literacy and/or numeracy in the same sample using measures that were both performance-based and self-reported. We excluded studies dealing with health literacy or numeracy of medical providers. Mirroring the outcomes studied in the AHRQ review, we included both health outcomes (accuracy of risk perception, health related skills, health behaviors, adherence, disease prevalence and severity, quality of life) and the use of health services (office and emergency department visits, preventive services, hospitalizations), but excluded health knowledge (for print literacy studies only, not numeracy studies), decision-making, and patient-provider communication (given that these latter two outcomes were felt to be moderators, and not on the causal pathway).<sup>22</sup>

Two reviewers (ESK/SLS or SCB/LAH) independently assessed abstracts identified from the MEDLINE search for inclusion, with full studies being retrieved if one or both reviewers selected an abstract for further review.

**Table 2. MEDLINE Search String**

Query	String
#1	Search numeracy
#2	Search "health literacy"
#3	Search #1 OR #2
#4	Search literacy
#5	Search "rapid estimate of adult literacy" OR real*
#6	Search #4 AND #5
#7	Search "test of functional health literacy" OR tofhl*
#8	Search #4 and #7
#9	Search "Hebrew health literacy test" OR HHLT
#10	Search #4 AND #9
#11	Search "medical achievement reading test" OR MART
#12	Search #4 and #11
#13	Search "newest vital sign" OR NVS
#14	Search #4 AND #13
#15	Search "short assessment of health literacy" OR SAHLSA
#16	Search #4 AND #15
#17	Search "wide range achievement test" OR WRAT
#18	Search #4 AND #17
#19	Search "nutritional literacy" OR "literacy assessment for diabetes" OR LAD OR SIL OR "single item numeracy screener" OR DAHL OR "demographic assessment" OR BEHKA OR "brief estimate" OR "diabetes numeracy" OR "medical data interpretation" OR "subjective numeracy" OR "numeracy test"
#20	Search #4 AND #19
#21	Search #6 OR #8 OR #10 OR #12 OR #14 OR #16 OR #18 OR #20
#22	#3 OR #21
#23	Search #22 Limits: Human, English
#24	Search #23 Limits: Editorial, Letter, Case Reports
#25	Search #23 NOT #24
#26	Search #25 AND "2010/02/01"[Pub Date] : "2014/12/09"[Pub Date]

**Table 3. Inclusion/Exclusion Criteria**

Population of interest	Patients of all ethnicities and ages (including healthy subjects and family caregivers)
Intervention	Measurement of health literacy/numeracy using self-reported measure
Comparator	Measurement of health literacy/numeracy using performance-based measure
Outcomes	Any relevant health outcomes (disease-specific outcomes, global health status, health related skills, health behaviors, adherence, disease prevalence and severity, quality of life, accuracy of risk perception), as well as use of health services (office and ED visits, preventive services, hospitalizations). Health knowledge was an accepted outcome only for health numeracy studies.
Time allowed for outcomes to appear	Any (including cross-sectional data)
Time searched	One year prior to latest AHRQ review (1 February 2010) to search date (9 December 2014)
Study designs allowed	RCTs, other clinical trials, case control, cohort studies, cross-sectional studies. No case reports or case series ( $n < 10$ ). Qualitative studies, reviews and validation studies were also excluded.

## Quality Assessment

We assessed included studies using quality criteria adopted from the AHRQ systematic evidence review (Appendix Table 1).<sup>22</sup> Two reviewers (ESK and SLS) independently rated each study as good, fair, or poor, based on an assessment of selection bias, measurement bias, confounding factors and sample size. Quality review focused specifically on the quality of the study as related to our specific research question, even if that question differed from the primary intent of the study. We arbitrated quality reviews only if the overall study rating or the rating of any individual quality criteria differed by two categories (i.e., poor versus good). Studies receiving a good or fair rating were included in the final analysis. Studies receiving a poor rating were excluded.

## Data Synthesis and Analysis

One reviewer (ESK) abstracted information from the studies into a summary table, and a second reviewer (SLS) checked the content for accuracy. We performed qualitative syntheses of the literature on the relationship of health literacy and numeracy and health outcomes and collaboratively synthesized results during our analysis. We paid particular attention to possible differences in the relationship of health literacy (and numeracy) and health outcomes, based on the purpose of measures (screening or describing) and underlying measurement construction (a psychometric versus skills-based approach). We contacted the corresponding author of one included study and two excluded studies (initially included, but later determined to be of poor quality related to our research question) to obtain additional data not included in published papers.<sup>24–26</sup>

## RESULTS

We identified two studies from the AHRQ systematic review for potential inclusion in our review.<sup>24,27</sup> We then reviewed 2,043 titles from our MEDLINE search for possible relevance. After this initial screen, two independent reviewers reviewed 969 abstracts and, subsequently, the full text of 276 papers. Of those, 214 were excluded because they did not have both a performance-based and self-reported measure, 41 had no original data, seven did not have an outcome of interest, six had an excluded study design, and one was unrelated to the review question. Seven studies from the MEDLINE search were retained for quality assessment.<sup>25,26,28–32</sup> We then hand-searched the reference lists from the included studies and a recent review of health literacy measures.<sup>2</sup> These yielded one additional study for inclusion.<sup>33</sup> Thus, a total of ten studies were identified and quality graded (Fig. 1).

Of the ten retained studies, four were rated as fair<sup>24,29,30,33</sup> and six were rated as poor for the purposes of this study.<sup>25–28,31,32</sup> The latter were excluded from further analysis. A poor quality rating usually resulted from lack of multivariate analyses adjusting for potential confounders of the relationships

between self-reported and performance-based health literacy or numeracy and health outcomes.

The characteristics of the included studies are summarized in Table 4. Three included studies measured health literacy; one measured health numeracy. Studies measuring health literacy used the S-TOFHLA and the REALM as performance-based measures. These were compared to the following self-reported measures: a question assessing confidence with medical forms adapted from Chew et al.'s work,<sup>8</sup> the Brief Health Literacy Screening Tool (BRIEF),<sup>12</sup> and an unvalidated question assessing self-reported problems with reading prescription labels. Sample sizes for these studies ranged from 100 to 378 patients, with all data collected representing convenience samples of patients in outpatient clinic settings. Outcomes were mostly self-reported, including patient-completed arthritis severity scores, self-reported diabetes, hypertension and stroke status, and patients' skill in interpreting their medication regimens. One study used an objective physician-completed arthritis severity score as an outcome measure.

The study comparing health numeracy measures<sup>24</sup> asked the following question adapted from the work of Lipkus et al.<sup>7</sup> to measure performance-based numeracy: "Which of the following numbers represents the biggest risk of getting a disease: 1 in 100, 1 in 1,000 or 1 in 10?" The study used the following question taken from the STAT-confidence scale<sup>16</sup> to assess self-reported numeracy: "In general, how easy or hard do you find it to understand medical statistics?" coded as very easy/easy or hard/very hard. This study included a nationally representative community sample with a large sample size (1,436 performance-based observations and 3,286 self-rated observations). The outcome was self-reported likelihood of colorectal cancer (CRC) screening and was assessed by mail or phone.

## Studies Examining Alternate Measures of Health Literacy

Three studies focused on the relationship between performance-based and self-reported measures of health literacy and health outcomes (Table 4). None focused on the comparison of performance-based measures and self-reported measures as their primary study question.

Haun and colleagues studied the relationship of the REALM, S-TOFHLA, and BRIEF on variables typically associated with low health literacy (e.g., age, education or disability), as well as several self-reported cardiovascular risk factors. Their study included 378 veterans at eight ambulatory VA clinics and found that significantly more patients were classified as having limited health literacy when assessed with the self-report measure, the BRIEF (57 % limited HL), than when assessed with either the REALM (37 %) or S-TOFHLA (17 %). However, they found no significant relationship between limited health literacy and three dichotomous health conditions (patient self-report of having or not having hypertension, diabetes, or a past stroke) using any of the health literacy measures after adjusting for age,

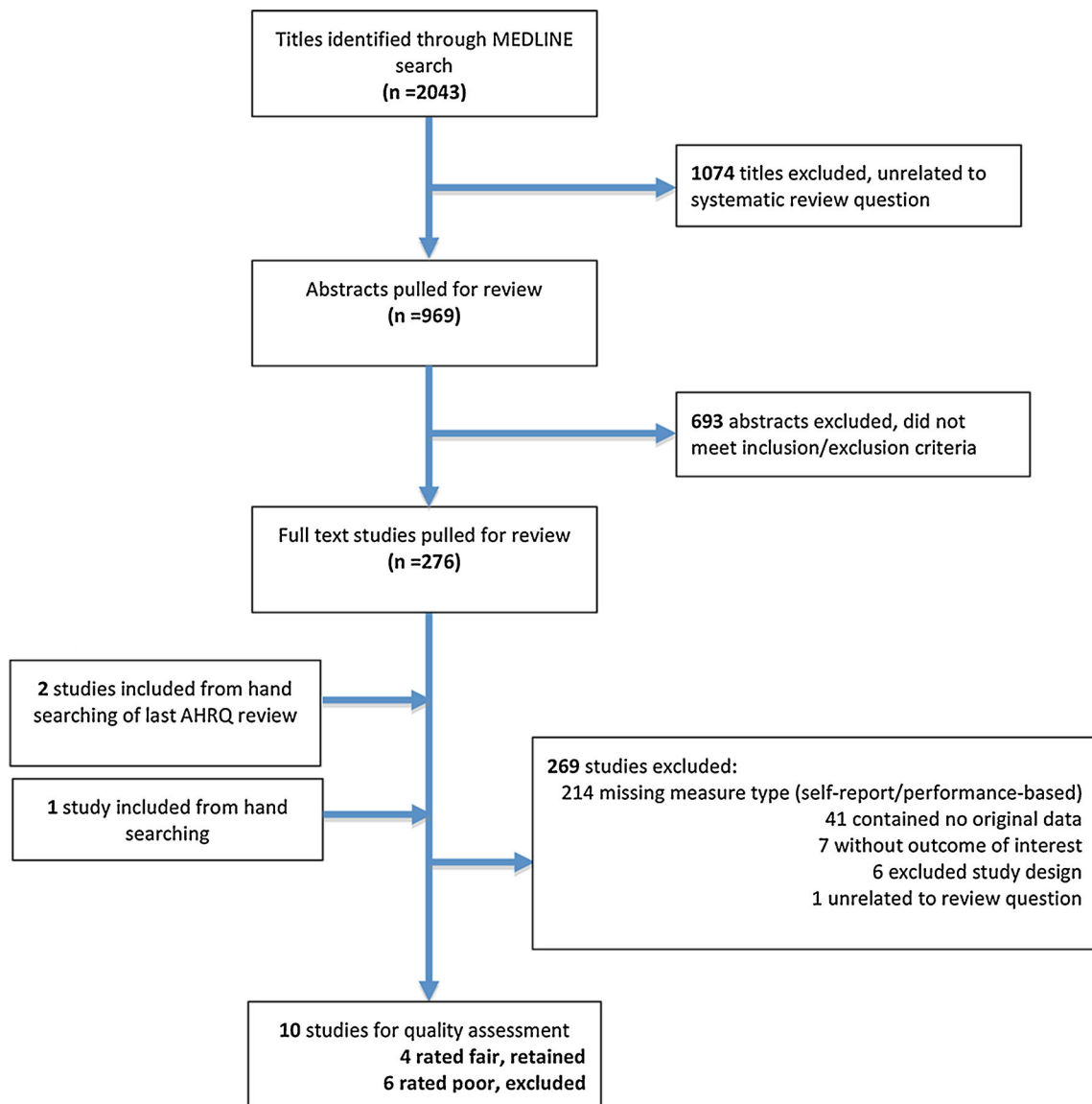


Figure 1 PRISMA flow diagram.

gender, race, education, self-reported reading level, retiree-status, and having a functional disability.

Hirsh and colleagues examined the relationship between the REALM, S-TOFHLA, and a single self-reported question assessing confidence with medical forms<sup>34</sup> and the outcome of rheumatoid arthritis severity. In 110 adults at a single rheumatology clinic, they found that 30 % of adults were deemed to have limited health literacy using the confidence question, compared to 49 % by the REALM and 35 % by the S-TOFHLA. Disease severity of patients' rheumatoid arthritis was assessed through both a physician-completed disease activity scale (DAS-28)<sup>35</sup> and a patient-completed tool (the Multidimensional Health Assessment Questionnaire, or MDHAQ).<sup>36</sup> In multivariate analyses adjusting for all significant variables in the study, they found that the patient-completed tool, the MDHAQ, was significantly associated with the confidence question. Specifically, each incremental

improvement in confidence, such as going from “not at all confident” to “quite a bit confident” was associated with a half point decrease in the MDHAQ score (range 0–3). However, no statistically significant associations were found between the REALM or the S-TOFHLA and the MDHAQ. None of the three literacy measures were significantly associated with the physician-completed scale, the DAS-28.

Marks and colleagues examined the relationship between patient demographics and measures of health literacy and medication knowledge and skill using the Medication Knowledge Score (MKS). This score asks patients to identify the names and dosages of medications from their pill bottle (a task we considered a skill based on criteria from the AHRQ review) and state the indications and potential side effects of their medications. In this study, Marks and colleagues found that a single self-reported question (unnamed) regarding ability to read medication labels identified 10 % of 99 patients as having some difficulty (7 %) or

Table 4 Characteristics of Included Studies

Study author (Quality rating) and study characteristics	Health literacy/numeracy measures and rates; and study outcome measures	Relationship between health literacy/numeracy and health outcomes												
<b>Studies measuring health literacy</b>														
<p>Haun<sup>30</sup> (fair)</p> <p><i>Purpose</i> To examine the variation in risk factors associated with health literacy across three instruments.</p> <p><i>Design</i> Cross-sectional in person survey</p> <p><i>Sample</i> 378 patients in ambulatory clinics in rural and non-rural VA medical facilities</p> <p><i>Demographics</i> Mean Age: 61.5 y 94 % Male 74 % White</p>	<p><i>Health Literacy Measures and Rates</i></p> <p><u>S-TOFHLA:</u> 17 % inadequate/marginal (score &lt; 23)</p> <p><u>REALM:</u> 37 % &lt; 9th grade (score&lt;61)</p> <p><u>BRIEF:</u> 57 % inadequate or marginal (score &lt;17)</p> <p><i>Outcome Measures</i> Diabetes, Hypertension and Stroke, each as a dichotomous health status indicator self-reported by the patient</p>	<p>Summary: No statistically significant differences in the relationship between limited health literacy and health outcomes for any comparisons after adjustment for age, sex, race, education, reading level, retired status, disability status. However, point estimates showed trends toward qualitative differences.</p> <p>Individual Findings: Adjusted Odds Ratio of inadequate/marginal health literacy among those with various outcomes*, 95 % CI</p> <table border="0"> <tr> <td><b>Diabetes:</b></td> <td><b>Hypertension:</b></td> <td><b>Stroke:</b></td> </tr> <tr> <td><u>S-TOFHLA:</u> 0.57, 0.28–1.16</td> <td><u>S-TOFHLA:</u> 1.10, 0.52–2.36</td> <td><u>S-TOFHLA:</u> 1.68, 0.71–3.98</td> </tr> <tr> <td><u>REALM:</u> 0.60, 0.35–1.02</td> <td><u>REALM:</u> 1.16, 0.67–2.00</td> <td><u>REALM:</u> 0.80, 0.36–1.80</td> </tr> <tr> <td><u>BRIEF:</u> 1.15, 0.70–1.87</td> <td><u>BRIEF:</u> 1.28, 0.78–2.11</td> <td><u>BRIEF:</u> 1.38, 0.63–3.00</td> </tr> </table>	<b>Diabetes:</b>	<b>Hypertension:</b>	<b>Stroke:</b>	<u>S-TOFHLA:</u> 0.57, 0.28–1.16	<u>S-TOFHLA:</u> 1.10, 0.52–2.36	<u>S-TOFHLA:</u> 1.68, 0.71–3.98	<u>REALM:</u> 0.60, 0.35–1.02	<u>REALM:</u> 1.16, 0.67–2.00	<u>REALM:</u> 0.80, 0.36–1.80	<u>BRIEF:</u> 1.15, 0.70–1.87	<u>BRIEF:</u> 1.28, 0.78–2.11	<u>BRIEF:</u> 1.38, 0.63–3.00
<b>Diabetes:</b>	<b>Hypertension:</b>	<b>Stroke:</b>												
<u>S-TOFHLA:</u> 0.57, 0.28–1.16	<u>S-TOFHLA:</u> 1.10, 0.52–2.36	<u>S-TOFHLA:</u> 1.68, 0.71–3.98												
<u>REALM:</u> 0.60, 0.35–1.02	<u>REALM:</u> 1.16, 0.67–2.00	<u>REALM:</u> 0.80, 0.36–1.80												
<u>BRIEF:</u> 1.15, 0.70–1.87	<u>BRIEF:</u> 1.28, 0.78–2.11	<u>BRIEF:</u> 1.38, 0.63–3.00												
<p>Hirsh<sup>29</sup> (fair)</p> <p><i>Purpose</i> To test the relationship between health literacy and rheumatoid arthritis</p> <p><i>Design</i> Cross-sectional in-person survey</p> <p><i>Sample</i> 110 adults in a rheumatology clinic at Denver Health</p> <p><i>Demographics</i> Mean Age: 53 y 21 % Male 27 % White</p>	<p><i>Health Literacy Measures and Rates</i></p> <p><u>S-TOFHLA:</u> 35 % inadequate/marginal (score &lt; 23)</p> <p><u>REALM:</u> 49 % &lt; 9th grade (score&lt;61)</p> <p><u>Brief Screening Question:</u> 30 % not at all/a little bit/somewhat confident filling out medical forms</p> <p><i>Outcome Measures</i> Multidimensional Health Assessment Questionnaire (MDHAQ), a patient completed scale assessing ten activities of daily living</p> <p>Disease Activity Score 28 (DAS-28), a physician completed rheumatoid arthritis severity scale.</p>	<p>Summary: Increase in self-reported health literacy on the brief screening question was associated with improvement in MDHAQ score after adjustment for age, sex, race, education, disease duration, marital status, tobacco, disease markers and treatment regimen. No significant association was found for the S-TOFHLA or REALM and MDHAQ. None of the measures were significantly associated with the DAS-28.</p> <p>Individual Findings: Adjusted Beta-Coefficients reflecting change in outcome by health literacy, 95 % CI</p> <p><b>MDHAQ (range 0–3):</b> <u>S-TOFHLA:</u> –0.010, –0.023 to 0.0024† <u>REALM:</u> –0.0067, –0.015 to 0.0014† <u>Brief Screening Question:</u> –0.50, –0.94 to –0.059 ‡</p> <p><b>DAS-28 (range 0–10)</b> <u>S-TOFHLA:</u> –0.016, –0.045 to 0.016† <u>REALM:</u> –0.0096, –0.030 to 0.011† <u>Brief Screening Question:</u> –0.47, –1.10 to 0.16†</p>												
<p>Marks<sup>33</sup> (fair)</p> <p><i>Purpose</i> To compare demographics and REALM scores and their prediction of medication knowledge/skill</p> <p><i>Design</i> Cross sectional in-person survey</p> <p><i>Sample</i> 100 patients seen at academic internal medicine clinic</p> <p><i>Demographics</i> Mean Age: 62 y 47 % Male 47 % White</p>	<p><i>Health Literacy Measures and Rates</i></p> <p><u>REALM:</u> 59 % &lt; 9th grade (score&lt;61)</p> <p><u>Brief Screening Question:</u> 10 % reporting some difficulty or total inability to read prescription label.</p> <p><i>Outcome Measure</i> Medication Knowledge Score, indicating the ability (or skill) to identify medications by pill bottle and the knowledge of name, dosage, indication and side effects of patients’ medications.</p>	<p>Summary: After adjusting for age and sex, REALM score was significantly related with the Medication Knowledge Score, while the brief screening question was not.</p> <p>Individual Findings: Adjusted Beta-Coefficient reflecting change in the outcome by health literacy, <i>p</i> values§</p> <p><b>Medication Knowledge Score (range 0–4):</b> <u>REALM:</u> 0.015 (&lt; 0.0001)</p> <p><u>Brief Screening Question:</u> Not reported because not statistically significant at <i>p</i>&lt;0.05  </p>												

(continued on next page)

Table 4. (continued)

Study author (Quality rating) and study characteristics	Health literacy/numeracy measures and rates; and study outcome measures	Relationship between health literacy/numeracy and health outcomes
<b>Studies measuring health numeracy</b>		
Ciampa <sup>24</sup> (fair)	<i>Numeracy Measures and Rates</i>	Summary: No statistically significant difference between Lipkus and STAT Confidence questions and colorectal cancer screening status after adjusting for age, race, education, income and insurance status. Low numeracy on either question was significantly associated with lower likelihood of being up to date on colorectal cancer screening.
<i>Purpose</i> To study the relationship between numeracy and perceptions of provider communication, as well as colorectal cancer screening utilization	<u>Single Lipkus Risk Question:</u> 22.6 % answered incorrectly  <u>STAT Confidence Question:</u> 39.4 % rating medical statistics hard/very hard to understand	Individual Findings: Adjusted Odds Ratio of answering numeracy question incorrectly if colorectal screening utilization*, 95 % CI
<i>Design</i> Cross-sectional mailed/phone based survey	<i>Outcome Measure</i> Colorectal cancer screening utilization, self-reported by patient as up-to-date or not (colonoscopy within 10 years, sigmoidoscopy within 5 years, or fecal occult blood test within 1 year)	<b>Colorectal Cancer Screening Utilization:</b> <u>Lipkus Question:</u> 0.61 (0.43–0.85) <u>STAT Confidence question:</u> 0.82 (0.68–0.98)
<i>Sample</i> National telephone and mailed survey with differential numeracy assessments by mode of delivery (1436 with both Lipkus and STAT confidence questions; 1850 had STAT confidence questions only)		
<i>Demographics</i> Mean Age: 63 y 47 % Male 78 % White		

\*Mathematically equivalent to RR (i.e., risk of outcome if limited health literacy) if the outcome is rare

†Unadjusted value. Adjusted value not reported because  $p > 0.10$

‡Association for responding “extremely” or “quite a bit” confident filling out medical forms

§Confidence intervals not reported

||In unadjusted analyses, mean MKS among those reporting some difficulty or total inability to read prescription labels 2.00, mean MKS among those reporting no difficulties reading prescription labels 2.43,  $p = 0.11$

being unable (3 %) to read medication labels, whereas 59 % of patients were classified as having inadequate or marginal health literacy on the REALM. In adjusted multivariate analysis, the REALM was associated with the MKS. By contrast, the result from the self-report question was not significantly associated with the MKS.

## Studies Examining Alternate Measures of Numeracy

One study focused on numeracy and its relationship with up-to-date status on CRC screening.<sup>24</sup> In this study, 22.6 % of a nationally representative sample of 1,436 patients contacted by mail and phone failed to correctly answer the performance-based numeracy question, while 39.4 % of patients reported that they found it hard or very hard to understand medical statistics.

In extra data obtained from the authors, both the performance-based and self-reported numeracy questions were found to be associated with CRC screening after adjusting for age, race, annual income, education and insurance status (odds ratio for up-to date CRC screening with low numeracy using performance-based measure: 0.61, 95 % CI 0.43–0.85; with low numeracy using self-reported question: 0.82, 95 % CI 0.68–0.98). However, in stratified analyses from the same study, performance-based and self-reported measures were differentially related to CRC screening utilization when stratified by several communication behaviors. Low *self-reported*

numeracy had no relationship with up-to-date CRC screening when patients reported that they always had a chance to ask health professionals all the health-related questions they had, or when they reported that their feelings and emotions were always given the attention they needed by health professionals. However, low *performance-based* numeracy was associated with lower up-to-date screening even when participants had the chance to ask questions and get the attention they needed. There was no difference in the relationship between low self-reported or performance-based numeracy and CRC screening utilization when participants were stratified by involvement in decision-making or by whether healthcare providers checked understanding of health-related information.

## DISCUSSION

Our systematic review highlights the paucity of literature regarding differences in the relationship of performance-based and self-reported measures of health literacy and numeracy with health outcomes. We identified only four relevant, fair quality studies, and none had a primary purpose of examining this relationship. These studies included a range of health literacy and numeracy measures with different purposes (screening versus description) and strategies for measurement construction (psychometric versus skills-based assessments). Additionally, each examined a range of health outcomes that

were often self-reported. The studies measuring the relationship between health literacy and outcomes found no differences in the relationship between performance-based and self-reported health literacy for four of six outcomes (self-reported diabetes, stroke, hypertension, and a physician-completed rheumatoid arthritis disease activity score). Health literacy measures were differentially related to a patient-completed health assessment questionnaire, and to a Medication Knowledge Score, although analyses were not adjusted for the same potential confounders. The study measuring the relationship between numeracy and health outcomes also found mixed results.

The few existing studies examining the relationship between performance-based and self-reported measures and health outcomes suggest a complex relationship. Furthermore, other studies that didn't meet our inclusion criteria for various reasons also found mixed results. In a letter to the editor, Daniel et al. reported that the S-TOFHLA was correlated with understanding hypothetical health care plans while a single item literacy screen was not, a difference that may arise from the comparison of a screening measure to a more comprehensive instrument.<sup>37</sup> In a validation study for the Subjective Numeracy Scale (SNS), Zikmund-Fisher et al. found that both performance-based numeracy and the SNS similarly predicted interpretation of numerical information.<sup>38</sup> Several studies that were excluded from our review because they did not adjust for the likely differences in baseline characteristics among those with self-reported and performance-based low literacy/numeracy (and were thus of "poor" quality for the purposes of this review) also suggest mixed results.<sup>25–28,31,32,39</sup>

An important question is why performance-based and self-reported measures may be differentially related to health outcomes. There are a few possible explanations. One explanation is that these measures are tapping into different latent constructs.<sup>2,20</sup> Performance-based measures often target skills such as reading comprehension, word recognition, and basic facility with numbers. Self-reported measures, on the other hand, may be tapping into something different. They generally assess a patient's perceived ability to perform a task, and may jointly assess confidence and social resources and skills, as well as pure print or numerical ability. Further, self-reported measures are less likely to undergo a full psychometric analysis. Another potential explanation may be differences in the purpose of the measure. Many self-reported measures are designed as screening tests, which may be differentially sensitive and specific than measures developed to more fully describe health literacy for research or clinical purposes. Further, performance-based and self-reported measures may interact differently with measures of cognition, a proposed driver of limited health literacy in certain populations.<sup>40–42</sup>

One final possible explanation is differentially aligned cutoffs for performance (Table 1). There is currently a lack of consensus on how high or low self-reported literacy and numeracy relate to performance-based cutoffs. In our review, some studies considered self-reported literacy or numeracy as binary screens, while others treated them as continuous variables. Many had no

"marginal" categorization similar to that in performance-based measures, or had not been previously validated. They were also tested in different populations, and their precision and reliability may be affected by these distinct environments. Such discrepancies may, in part, be responsible for different conclusions about the relationships between various measures and outcomes.<sup>1</sup>

To move the field forward, further studies are needed that directly compare multiple validated self-reported measures of health literacy and numeracy against a variety of objectively measured health outcomes in a single sample. The studies should pay particular attention to issues of underlying purpose and psychometric construction, and thereby compare single-item, self-reported literacy screens with single-item, performance-based literacy screens; and multi-item, self-reported scales with multi-item, performance-based measures. Studies should also pick aligned cut-points prior to examination of the relationship of health literacy and numeracy with health outcomes.

In considering this work, readers should consider limitations of our review and the available literature. Beyond the general limitations of the available literature, all included studies were cross-sectional in design, making it impossible to discuss the role of causality in the associations found between health literacy or numeracy and outcomes. Randomized controlled trials, or other prospective study designs, could more accurately describe the relationship between the two. Another limitation is selection bias within the included studies; we expect that low literacy patients, particularly those with fewer resources, may have declined study participation for fear of embarrassment or shame, a concern reported by other studies.<sup>18,43</sup> Had these patients participated, studies may have yielded different results. Further, in following the 2011 AHRQ protocol for relevant outcomes, we did not include studies examining the relationship of health literacy and knowledge. These studies may be available and could be examined at a future time. Additionally, we included the medication knowledge score as a skill-based outcome, but acknowledge that this outcome overlaps with performance-based measurement of health literacy in the S-TOFHLA and other literacy measures, making interpretation challenging. Continued discussions about the relationships between literacy, numeracy, and skills-based outcomes will be important moving forward. Finally, these results may not generalize across other health conditions or health outcomes.

## CONCLUSION

We found a paucity of studies examining the relationship between performance-based and self-reported measures of health literacy and numeracy and health outcomes, and no studies designed specifically to address this question. The results of available studies were mixed. To further understand whether performance-based and self-reported measures are differentially related to health outcomes, future studies should assess multiple performance-based and self-reported measures in a single sample, and use objective measures of health outcomes.



**Conflicts of Interest:** Stacy Bailey has worked as a consultant for Merck, Sharp, Dohme and MedThink SciCom. She has received grants from Merck, Sharp and Dohme, and worked as a co-investigator on grants and contracts from United HealthCare and Abbott. Laurie Hedlund has worked as a consultant for Luto. All other authors declare that they do not have a conflict of interest.

**Corresponding Author:** Sheridan, MD MPH; University of North Carolina at Chapel Hill, 5039 Old Clinic Building, CB #7110, Chapel Hill, NC 27599-7110, USA (e-mail: stacey\_sheridan@med.unc.edu).

## REFERENCES

- Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med.* 2011;155(2):97.
- Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol.* 2011;64(4):366–79.
- McCormack L, Haun J, Sørensen K, Valerio M. Recommendations for advancing health literacy measurement. *J Health Commun.* 2013;18(sup1):9–14.
- Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999;38:33–42.
- Davis TC, Long SW, Jackson RH, Mayeaux E, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med.* 1993;25(6):391.
- Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127(11):966–72.
- Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Mak.* 2001;21(1):37–44.
- Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med.* 2004;36(8):588–94.
- 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(1):21.
- Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. *J Med Internet Res.* 2006;8(4):e27.
- Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. *Diabetes Care.* 2008;31(5):874–9.
- Haun J, Dodd V, Varnes J, Graham-Pole J, Rienzo B, Donaldson P. Testing the brief health literacy screening tool: implications for utilization of a BRIEF health literacy indicator. *Fed Pract.* 2009;26(12):24–8.
- Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Mak.* 2007;27(5):672–80.
- Chinn D, McCarthy C. All Aspects of Health Literacy Scale (AAHLS): Developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. *Patient Educ Couns.* 2013;90(2):247–53.
- Jordan JE, Buchbinder R, Briggs AM, Elsworth GR, Busija L, Batterham R, et al. The Health Literacy Management Scale (HeLMS): A measure of an individual's capacity to seek, understand and use health information within the healthcare setting. *Patient Educ Couns.* 2013;91(2):228–35.
- Woloshin S, Schwartz LM, Welch HG. Patients and medical statistics. *J Gen Intern Med.* 2005;20(11):996–1000.
- Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int.* 2000;15(3):259–67.
- Wolf MS, Williams MV, Parker RM, Parikh NS, Nowlan AW, Baker DW. Patients' shame and attitudes toward discussing the results of literacy screening. *J Health Commun.* 2007;12(8):721–32.
- Ferguson B, Lowman SG, DeWalt DA. Assessing literacy in clinical and community settings: the patient perspective. *J Health Commun.* 2011;16(2):124–34.
- Haun JN, Valerio MA, McCormack LA, Sørensen K, Paasche-Orlow MK. Health literacy measurement: an inventory and descriptive summary of 51 instruments. *J Health Commun.* 2014;19(sup2):302–33.
- Pleasant A. Advancing health literacy measurement: A pathway to better health and health system performance. *J Health Commun.* 2014;19(12):1481–96.
- Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Viera A, Crotty K, et al. Health literacy interventions and outcomes: an updated systematic review. Evidence Report/Technology Assessment. United States: Agency for Healthcare Research and Quality; 2011 Mar. Report No.: 199.
- DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes. *J Gen Intern Med.* 2004;19(12):1228–39.
- Ciampa PJ, Osborn CY, Peterson NB, Rothman RL. Patient numeracy, perceptions of provider communication, and colorectal cancer screening utilization. *J Health Commun.* 2010;15(S3):157–68.
- Morris NS, Field TS, Wagner JL, Cutrona SL, Roblin DW, Gaglio B, et al. The association between health literacy and cancer-related attitudes, behaviors, and knowledge. *J Health Commun.* 2013;18(sup1):223–41.
- Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. *Psychol Aging.* 2013;28(4):1124.
- Sheridan SL, Pignone MP, Lewis CL. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *J Gen Intern Med.* 2003;18(11):884–92.
- Briggs AM, Jordan JE, O'Sullivan PB, Buchbinder R, Burnett AF, Osborne RH, et al. Individuals with chronic low back pain have greater difficulty in engaging in positive lifestyle behaviours than those without back pain: An assessment of health literacy. *BMC Musculoskelet Disord.* 2011;12(1):161.
- Hirsh JM, Boyle DJ, Collier DH, Oxenfeld AJ, Nash A, Quinzanos I, et al. Limited health literacy is a common finding in a public health hospital's rheumatology clinic and is predictive of disease severity. *J Clin Rheumatol.* 2011;17(5):236–41.
- Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. *J Health Commun.* 2012;17(Suppl 3):141–59.
- Rakow T, Wright RJ, Bull C, Spiegelhalter DJ. Simple and multistate survival curves: can people learn to use them? *Med Decis Making.* 2012;32(6):792–804.
- Koay K, Schofield P, Gough K, Buchbinder R, Rischin D, Ball D, et al. Suboptimal health literacy in patients with lung cancer or head and neck cancer. *Support Care Cancer.* 2013;21(8):2237–45.
- Marks JR, Schectman JM, Groninger H, Plews-Ogan ML. The association of health literacy and socio-demographic factors with medication knowledge. *Patient Educ Couns.* 2010;78(3):372–6.
- Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, 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.
- Prevoo M, Van't Hof M, Kuper H, Van Leeuwen M, Van de Putte L, Van Riel P. Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* 1995;38(1):44–8.
- Pincus T, Sokka T, Kautianinen H. Further development of a physical function scale on a multidimensional health assessment questionnaire for standard care of patients with rheumatic diseases. *J Rheumatol.* 2005;32(8):1432–9.
- Daniel D, Greene J, Peters E. Screening question to identify patients with limited health literacy not enough. *Fam Med.* 2010;42(1):7–8.
- Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitation. *Med Decis Mak.* 2007;27(5):663–71.
- Sheridan SL, Pignone MP. Numeracy and the medical student's ability to interpret data. *Eff Clin Pract.* 2002;5(1):35–40.
- Wolf MS, Curtis LM, Wilson EA, Revelle W, Waite KR, Smith SG, et al. Literacy, Cognitive Function, and Health: Results of the LitCog Study. *J Gen Intern Med.* 2012;27(10):1300–7.
- Kaphingst KA, Goodman MS, MacMillan WD, Carpenter CR, Griffey RT. Effect of cognitive dysfunction on the relationship between age and health literacy. *Patient Educ Couns.* 2014;95(2):218–25.
- Ownby RL, Acevedo A, Waldrop-Valverde D, Jacobs RJ, Caballero J. Abilities, skills and knowledge in measures of health literacy. *Patient Educ Couns.* 2014;95(2):211–7.
- Parikh NS, Parker RM, Nurss JR, Baker DW, Williams MV. Shame and health literacy: the unspoken connection. *Patient Educ Couns.* 1996;27(1):33–9.