Quantitative Literacy on the Web of Science, 2 – Mining the Health Numeracy Literature for Assessment Items

H L. Vacher
Department of Geology, University of South Florida, vacher@usf.edu
Todd Chavez
University of South Florida, tchavez@usf.edu

Quantitative Literacy on the Web of Science, 2 – Mining the Health Numeracy Literature for Assessment Items

Abstract

A topic search of the Web of Science (WoS) database using the term "numeracy" produced a bibliography of 293 articles, reviews and editorial commentaries (Oct 2008). The citation graph of the bibliography clearly identifies five benchmark papers (1995-2001), four of which developed numeracy assessment instruments. Starting with the 80 papers that cite these benchmarks, we identified a set of 25 papers (1995-2008) in which the medical research community reports the development and/or application of health-numeracy assessments. In all we found 10 assessment instruments from which we have compiled a total of 48 assessment items. There are both general and context-specific tests, with the wide range in the latter illustrated by names such as the Diabetes Numeracy Test and the Asthma Numeracy Questionnaire. There is also a Medical Data Interpretation Test and a Subjective Numeracy Scale. Much of this literature discusses the validity and reliability of the test, and many papers include item-by-item results of the tests from when they were applied in the research reported in the papers. The research that used the tests was directed at exploring such subjects as the patients' ability to evaluate risks and benefits in order to make informed decisions; to understand and carry out instructions in order to self-manage their medical conditions; and, in research settings, to understand what the researchers were asking in their assessments (e.g., quantified quality of life) that require comparison of numerical information. We present the collection of items as a potential resource for educators interested in numeracy assessments in context.

Creative Commons License

www.manaraa.com

Introduction

In the previous issue of this journal, we presented the results of a keyword search of the *Web of Science* (*WoS*)[1] database (search date 1 March 2008) for numeracy, quantitative literacy and statistical literacy (Vacher and Chavez, 2008). That search produced a bibliography consisting of 338 records.[2] About 32% of the bibliography (109 records) was from the field of medical and health sciences. Most of those papers (70) were linked by citations (Fig. 1). The terms *health literacy* and *health numeracy* were conspicuous in the collection, and so was the use of assessment instruments.

The purpose of the study reported here was to explore this collection of health numeracy papers in the *WoS* database. We looked specifically for answers to two types of questions. First, what do users of the term "health numeracy" mean by "numeracy?" Second, what questions do they use to assess numeracy? Our perspective is entirely from the outside looking in at the health numeracy literature. We were interested in definitions and assessment items to take beyond the field of medical and health sciences to the broader numeracy community. In the process, we became fascinated to read how the assessment instruments are being used in research in the medical sciences and how these scientists established the reliability and validity of the tests.

Search Methodology

The topic search of our earlier study used the multiple disjunction "quantitative literacy" OR "numeracy" OR "statistical literacy." We learned in that project that the term of choice for papers in the health and medical sciences is *numeracy*; therefore, we simply used the term "numeracy" for the topic search of this study. That search (most recently 29 October 2008) produced a total of 293 articles, reviews and editorials.[3] We exported them to HistCite[4] to mine for papers involving numeracy assessments.

We refined the 293-paper numeracy bibliography in a variety of ways. First we used the word tool in HistCite to search for the following keywords: "assessing" (producing 9 papers), "assessment" (5), "test" (13) and "testing" (1). These four refinements produced a total of 25 unique references, 15 of which we judged from the titles might include or have used assessment items. From reading

---

[1] http://www.thomsonreuters.com/products_services/scientific/Web_of_Science, accessed 12/29/08
[2] http://www.lib.usf.edu/Numeracy/V-C_2008-07/, accessed 12/29/08: 338 records consisting of articles (301), reviews (16), editorial material (18), and notes (3).
[3] Latest before press: http://www.lib.usf.edu/Numeracy/V-C_2008-12/. Created 12/29/08: 303 records consisting of articles (268), reviews (20), editorial material (13), and notes (2)
[4] http://www.histcite.com/, accessed 12/29/08

the 15 selected references, we found some studies that indeed developed and/or applied assessment instruments that were included in the paper, but it also became evident that there were earlier assessments of health numeracy in the 293-paper numeracy bibliography. From the citations, we added the earlier papers to our reference list and compared it to the HistCite-generated citation graph of the numeracy bibliography. Five of the six most evident hub papers in the citation graph (Fig. 1) were on our reference list. We used those five papers in a second refinement of the 293-paper numeracy bibliography: using HistCite, we found the papers in the numeracy bibliography that cite the five benchmarks:

- Williams et al. (1995): 13 papers
- Schwartz et al. (1997): 49 papers
- Baker et al. (1999): 12 papers
- Lipkus et al. (2001): 41 papers
- Woloshin et al.(2001): 13 papers

This refinement produced a list of 80 references that cited one or more of the early health-numeracy assessments papers.
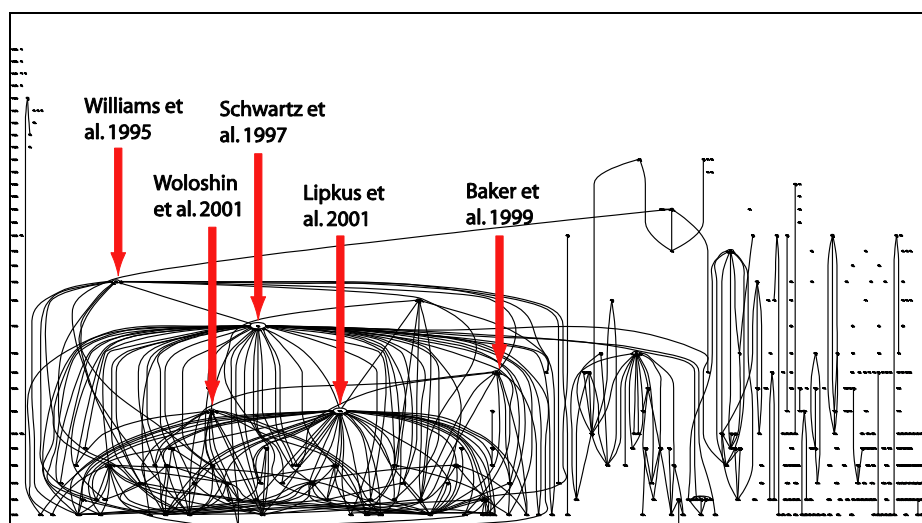
Figure 1. Citation graph of the numeracy-quantitative literacy-statistical literacy bibliography, July 2008, with the five benchmark papers used in this paper to search the *WoS* health numeracy bibliography (adapted from Vacher and Chavez 2008).

Reading the abstracts or, where necessary, browsing the full text of the 80 references and chasing down new references cited in them produced a final list for further consideration in our survey of health numeracy assessment items. In the end, our mining operation produced 25 papers that developed and/or applied

health numeracy assessment items to measure the numeracy of subjects (generally patients). In all, we found ten assessment instruments.

Pushing the mining analogy, one can say that the numeracy bibliography of 293 papers was the ore deposit. The first 15-item list was an exploratory bore hole that allowed us to locate health numeracy-assessment ore. We dug out ore amongst the 80 references, separating 25 ore papers from 55 papers that went into tailings. We sent the 25 ore papers to the mill to extract assessment items.

Extending the mining analogy further, we refer to the structure of a metalliferous ore deposit (e.g., copper, lead, zinc) of the kind that occurs in geologically young mountain provinces (e.g., western North and South America). The so-called primary[5] (i.e., early formed) ore lies deep underground having been formed in association with igneous intrusions that accompanied the mountain building; the primary ore features massive, dense sulfide minerals. Closer to the surface is the secondary, weathered zone where the primary minerals have reacted with oxygenated, meteoric (atmosphere-derived) groundwater charged with carbon dioxide. At places, exquisite museum-quality crystals of secondary carbonate and sulfate ore minerals line cavities and vugs in the upper zone of the ore deposit. Miners often collect such specimens for mineral displays.

With such an ore deposit in mind, we organize our presentation as follows. First we introduce some of the papers of the secondary zone—e.g., what is health numeracy?—and display some of the good specimens we collected (our "mineral crystals"). Then we come to the massive, dense literature of the primary ore, which contains the assessment instruments we sought in association with the research by which and for which they were developed. We end with a description of the extracted product, the assessment items.

Definitions and Commentary (SECONDARY ORE MINERALS)

What is Health Literacy?

The paper by Speros (1995), "Health literacy: Concept analysis," can serve to introduce the subject. The author writes (p. 635):

> The concept of health literacy is used in health science literature and discussions as a variable that relates to health outcomes. Health literacy level has emerged as an outcome related to adherence to prescribed health care recommendations and patient empowerment….

---

[5] From Bateman (1950, p. 20): "Ore minerals are also classed as primary … and secondary…. The former were deposited during the original period or periods of metallization; the latter are alteration products of the former as a result of weathering or other surficial processes resulting from descending surface waters.

Continuing, Speros gives three definitions as prominent ones in the literature (p. 635):

- "a constellation of skills, including the ability to perform basic reading and numerical tasks required to function in the health care environment" (American Medical Association Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs 1999, p. 553)
- "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" (U.S. Department of Health and Human Services 2000 p. 11−20.
- "… the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways that promote and maintain good health" (World Health Organization 1998, p. 10).

Under "Defining attributes," she lists the following five as ones that appear consistently in the literature: (1) reading skills, (2) numeracy skills, (3) comprehension, (4) capacity to use information in health care decision-making, and (5) successful functioning in the role of health care consumer. Regarding the second attribute, she writes (p. 636):

> The ability to read and understand numbers and perform basic mathematical computations is referred to as 'numeracy' skill.

She goes on to say numeracy is defined by experts as "the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed material" and cites the Human Resources Development Canada (HRDC 1997, p. 14) for the definition. In the next sentence she references a publication of the National Academies Press (Nielson-Bohlman et al. 2004) for the statement that "numeracy, or quantitative literacy, may be the most important element of health literacy."

The references from the AMA Committee, Health and Human Services, WHO, HRDC, and Nielson Bohlman et al. are not indexed by the *Web of Science* and, therefore, are not included in our 293-paper numeracy bibliography. Other broad treatments of health literacy that are in the bibliography include "Health literacy: Implications for family medicine" (Davis and Wolf 2004), "Understanding health literacy: an expanded model" (Zarcadoolas et al. 2005), and "The meaning and measure of health literacy" (Baker 2006).

What is Health Numeracy?

One of the papers in our bibliography is an editorial by Montori and Rothman (2005), "Weakness in numbers: the challenge of numeracy in health care." It

4

includes the following short statement connecting health literacy and numeracy (p. 1071):

> Health literacy is the ability to complete basic reading and numerical tasks required to perform in the health care environment. The specific aspect of literacy that involves solving problems requiring understanding and use of quantitative information is sometimes called numeracy. Numeracy skills include understanding basic calculations, time and money, measurement, estimation, logic and performing multistep operations. Most importantly, numeracy also involves the ability to infer what mathematical concepts need to be applied when interpreting specific situations, and to use this information to problem solve.

The statement does not contain the term health numeracy, although it would not be difficult to construct a definition from those four sentences.

The paper by Golbeck et al. (2005), "A definition and operational framework for health numeracy," contains the definition that is quoted in many papers (e.g., Donelle et al. 2007; Ancker and Kaufman 2007; Lu et al. 2008; Nelson et al. 2008):

> Health numeracy is the degree to which individuals have the capacity to access, process, interpret, communicate, and act on numerical, quantitative, graphical, biostatistical, and probabilistic health information needed to make effective health decisions.

The authors then go on to define four, somewhat overlapping categories of health numeracy:

- *Basic health numeracy*, involving the "sufficient basic skills to identify numbers, and make sense of quantitative data requiring no manipulation of numbers." Their examples include using a phone book and identifying the time and date of an appointment.

- *Computational health numeracy*, involving "the ability to count, quantify, compute and otherwise use simple manipulation of numbers, quantities, items or visual elements in a health context so as to function in everyday health situations." Their examples include determining net carbohydrates from a nutritional label.

- *Analytical health numeracy*, a higher-level literacy, involving "concepts such as inference, estimation, proportions, percentages, frequencies, and equivalent situations" as well as pulling information from multiple sources and in multiple formats. Their examples include determining whether cholesterol levels are within the normal range and understanding basic graphs.

- *Statistical health numeracy*, involving "probability statements, skills to compare information presented on different scales (probability, proportion, percent), the ability to critically analyze quantitative health information such as life expectancy and risk, and an understanding of statistical concepts such a randomization and "blind" study." Their

examples involve working with probabilities of treatment efficacy and side effects, complex graphs, and relative vs. absolute risk.

In "A framework for health numeracy: How patients use quantitative skills in health care," Schapira et al. (2008) took an empirical approach. They analyzed six audio- and videotaped focus groups in which a total of 50 participants randomly selected from three primary care facilities associated with an academic medical center discussed how numbers are used in the health care setting. Their research resulted in a triangular hierarchical health-numeracy framework. The framework consists of three domains: primary numeracy skills at the base; applied health numeracy at an intermediate level; and interpretive health numeracy at the apex. Definitions and subcategories of these domains are listed in Table 1.

Table 1. Domains and subcategories of health numeracy (reformatted from Table 2 of Schapira et al. 2008)

Primary numeric skills: Ability to use basic arithmetic functions and graphs as well as apply numbers to concept of dates and time—Counting; basic math functions; estimating and projecting; dates and time; scales and graphs.
Applied numeracy • Basic tasks: The use of numbers in day-to-day health care tasks such as taking medications as prescribed—Interpretation of lab values; medication adherence; symptom scale (pain); scheduling appointments; paying bills. • Risk communication tasks: The use of numbers to communicate probabilistic information about health outcomes including risk, severity, and outcomes of disease—Disease incidence; modification of incidence by risk factors or health behaviors; prognosis, survival; adverse outcome of intervention; efficacy of intervention; results of diagnostic tests; measures of disease severity. • Decision-making tasks: The use of numbers to help consider the risks and benefits of a given medical decision: Information seeking; balancing risks and benefits; assessment of evidence.
Interpretive numeracy: The ability to understand the strengths and limitations of numbers to represent health or disease states, the efficacy of an intervention, or other expected health outcomes—Probability and chance; principles of scientific method; concept of uncertainty; representative nature of numbers; graphic and verbal formats; individual and biologic variation in expected outcomes.

Ancker and Kaufman (2007) took an informatics approach. Their research started with keyword searches ("health literacy" and "numeracy") of MEDLINE, CINAHL, and PsychINFO databases and built on results from a prior study (Ancker et al. 2006) on how graphs are used to communicate health risk. The goal was to use the literature to categorize quantitative skills and characteristics that contribute to patients' "*productive use of quantitative health information*, i.e., the effective use of quantitative information to guide health behavior and make health decisions" (p. 713, emphasis authors'). They state (p. 713-714):

> Productive information use (for example, a patient's successful completion of a medication regimen) depends only in part on health numeracy, the individual-level skills to obtain, interpret, and process quantitative information….It also depends upon the ability of the expert or the information artifact to provide appropriate and cognitively manageable information. Even consumers with advanced math skills may perform poorly when trying to use poorly explained information; conversely, good design of information artifacts can compensate for weak individual-level skills. Thus, the productive use of the information can be considered a result of the entire system of health communication, not solely of the individual patient's skills.

With respect to the individual-level skills, Ancker and Kaufman identified three factors that contribute to patients' productive use of quantitative health information (Table 2).

Table 2. Factors contributing to a patient's ability to use quantitative information for health: individual-level skills (reformatted from Table 1 of Ancker and Kaufman 2007)

Patients' quantitative skills   Definition: Basic computational skills (e.g., addition, multiplication, and use of simple formulas), estimation, and statistical literacy.   Examples: Computing calorie content; comparing a computation to an estimate to determine whether it is correct; understanding concept of randomization in a clinical trial.
Patients' ability to use information artifacts   Definition: Ability to navigate documents, interpret graphs, and translate between different representations of the same information   Examples: Obtaining nutrient information from a nutrition label; comparing personal health data as displayed on different meters or devices.
Patients' oral communication skills   Definition: Ability to speak clearly about quantities and understand spoken information.   Examples: Reporting a previous medication regimen accurately to a new physician.

Notably, the authors reached beyond the numeracy of individual patients and took the perspective of distributed cognition embracing the entire system of health communication. Their study included health numeracy papers focusing, for example, on the effects of changing representation (e.g., the design of graphs; the format of numbers) and the skills of the health-care provider. Thus they went beyond the three individual-numeracy skills in their table of factors affecting patients' ability to use quantitative information for health; they added five more (Table 3). These factors illustrate that the field of health numeracy also intersects with information design and the quantitative literacy of providers.

Where is the Cutting Edge?

That much is known about health numeracy is evidenced by a big review paper, "Clinical Implications of Numeracy: Theory and Practice" (Nelson et al. 2008) and its 128 references. For example (p. 271):

Low numeracy is pervasive and constrains informed patient choice, reduces medication compliance, limits access to treatments, impairs risk communication, and affects medical outcomes. Numeracy explains unique variance in medical decision making beyond that explained by such factors as education or intelligence, and cannot be reliably inferred by observable patient characteristics. Well-validated objective numeracy measures provide the most accurate assessment of basic numerical skill. Subjective self-report measures, which are correlated with objective numeracy measures, may be useful tools in the clinical context because patients find them less burdensome and less intimidating than objective measures….

Table 3. Factors contributing to a patient's ability to use quantitative information for health: Beyond individual-level skills (reformatted from Table 1 of Ancker and Kaufman 2007)

Information design for patients
Definition: Arrangement of information media and symbols to support comprehension and cognition.
Examples: Designing a patient interface for an electronic health record that provides graphics to illustrate numerical information.
Providers' oral communication skills
Definition: Ability to communicate quantitative concepts clearly to the patient.
Examples: Explaining a new medication regimen to a patient in an understandable fashion.
Providers' quantitative skills
Definition: Basic computational skills, estimation, and statistical literacy.
Examples: Converting between units of measure; understanding the positive predictive power of a diagnostic test.
Providers' ability to use information artifacts
Definition: Ability to navigate documents, interpret graphs, translate between representations of the same information.
Examples: Interpreting a graph of patient lab values over time; applying the numerical output of a decision support system to an individual case.
Information design for providers
Definition: Ability of a system or document to support the provider's cognition.
Examples: Designing a provider interface that provides automated conversions between units of measure.

Even so, in the section titled "Future Research in Health Numeracy," Nelson et al. say that the construct of health numeracy still wants for a definition:

At a conceptual level, there is a need for a consensus regarding a definition of health numeracy that is empirically and theoretically derived. By empirically, we mean a construct whose components are justified by empirical evidence, such as results of psychometric analysis. By theoretically, we mean a comprehensive theoretical account of health numeracy that can be used to make specific predictions.

The theoretical framework is needed too, they say, so that the construct can be fully operationalized. As it is, "the most widely used measures to assess health numeracy tap different dimensions of numeracy" (Nelson et al. 2008, p. 270).

In addition to the definition and measurement issues, Nelson et al. point to the area of communicating risk through visual and verbal means. What graphical displays are effective? Verbally, is risk communication clearer with numbers (e.g., 0.001% chance of a side effect) or with a qualitative expression (e.g., "a rare side effect"). Do the answers vary with numeracy?

The issue of finding what works for effective risk communication comes up repeatedly in editorials and summary accounts. The subject seems to have been energized by the recognition from the numeracy assessments that the level of numeracy in the patient population is poor (see next section). In "Contingent or universal approaches to *patient deficiencies* in health numeracy" (emphasis added), for example, Hamm and Bard (2007) write, "It seems reasonable to figure out how physicians and health communicators may compensate for patients' numeracy deficits by communicating in a way so that they may understand quantitative concepts." In a summary in a publication of project HOPE, Peters et al. (2007, p. 741−742) set the larger context:

> Increasingly, the emphasis of health care policy is to tap the potential power of informed consumers to improve health outcomes and the efficacy of health care. Employers and payers, recognizing the essential role that consumers can play in containing costs and improving care, have undertaken initiatives to influence consumers' behavior. They have urged consumers to change the way they select and use health care and how they manage their day-to-day health. Attempts have been made to encourage consumers to select high-performing providers, health plans, and facilities; choose evidence-based, cost-effective treatments; collaborate with their providers; initiate and maintain healthy behavior; and manage their own symptoms and conditions. At the same time, consumers are being asked to assume a greater share of their health care costs than ever before. As a result, choices have become more consequential for patients, in terms of both financial and health outcomes.
>
> However, not all consumers have the skills needed to use health information. A key concern is whether the policy approach of giving greater responsibility to patients will further disadvantage those with limited skills.

So what is the mathematics that causes so much difficulty? Reyna and Brainard (2007, p. 147) answer:

> Recent research on numeracy in medical decision making has shown that many adults fail to solve simple ratio and decimal problems, concepts that are prerequisites for understanding health-relevant risk communications. Along with research in education and cognitive development, this work demonstrates that adults have difficulty with a broad range of ratio concepts (including fractions, proportions, and probability judgments). Research confirms that this difficulty, as measured by content-neutral numeracy tests, predicts poorer health outcomes, less accurate perception of health risks, and a compromised ability to make medical decisions.

The finding that people have difficulties with ratios will resonate with many educators in the broader numeracy community, of course (e.g., Schield 2008;

Tucker 2008[6]). In a companion paper, "Numeracy, ratio bias, and denominator neglect in judgments of risk and probability," the authors (Reyna and Brainard 2008) give an extensive analysis of psychological processes that underlie the difficulties.

Assessments in Context (PRIMARY ORE)

Our search found five kinds of numeracy assessment instruments in the bibliography. The TOFHLA group (Williams et al. 1995, Baker et al., Gong et al. 2007) aims at functional health literacy (Nelson et al 2008, p. 264: "the ability to understand and act on numerical health information and thereby function effectively in the health care environment"); these three instruments test both reading comprehension and numeracy.   Two tests assess basic arithmetic skills (mainly ratios); the first (Schwartz et al. 1997) is content-free; the second (Lipkus et al. 2001), which builds on the first, adds questions that are focused on general health risks (i.e., getting an unspecified disease or infection). Three instruments focus on specific health conditions: anticoagulation control[7] (Estrada et al. 2004); asthma (Apter et al. 2006); diabetes (Huizinga et al. 2008).   One test (Schwartz et al. 2005) assesses patients' abilities to interpret and even combine numerical risk information.   Lastly, in contrast to all the others which are objective assessments, one (Fagerlin et al. 2007) is a subjective test; it asks the patients for self-assessments of their numerical ability and for their preferences on how numerical information is presented.

In the following, we highlight the assessment items in the ten assessment instruments and, for context, illustrate how they have been used in various research papers.   We take the instruments in chronological order.   We also summarize what we can of the information on reliability and validity contained in the papers in the bibliography[8].

TOFHLA – Test of Functional Health Literacy in Adults (Williams et al., 1995)

The vanguard paper[9] linking numeracy and health literacy opens with:

---

[6] From Steen (2008, p. 20), "Whereas Schield worries that students do not know what a fraction *means*, Tucker worries that they do not know what it i*s*."

[7] For patients taking warfarin (brand name, Coumadin), a blood thinner.

[8] See http://www.intestcom.org/Orta/Classical+test+theory.php (accessed 12/29/08) on the Web site of the International Test Commission for a convenient, open-access account that explains much of the terminology of psychometric testing.

[9] The paper by Williams et al. (1995) is the earliest by nearly four years of the 29 papers retrieved (23 October 2008) from *WoS* using the conjunction "health literacy" AND "numeracy." For a topic search for "health literacy" alone, it was 970th on the list of 975, meaning that there were five older papers.

> An estimated 40 million to 44 million adults in the United States are functionally illiterate, i.e., they cannot perform the basic reading tasks required to function in society according to the National Adult Literacy Survey. Another 50 million adults are only marginally literate. Among those with the lowest literacy skills, nearly half live in poverty, and one fourth report physical, mental, or health conditions that prevent them from participating fully in work, school, or housework.
>
> Lack of adequate literacy skills may be an important barrier to receiving proper health care. Patients are routinely expected to read and understand labels on medicine containers, appointment slips, informed consent documents, and health education materials. Anecdotal reports have described the difficulties encountered by illiterate patients, and several studies have shown health education materials and consent forms are often written at levels exceeding patients' reading level.

The paper, which has the title "Inadequate Functional Health Literacy among Patients at Two Public Hospitals," draws the following conclusion:

> Many patients at our institutions cannot perform the basic reading tasks required to function in the health care environment. Inadequate health literacy may be an important barrier to patients' understanding of their diagnoses and treatments, and to receiving high-quality care.

En route to that conclusion, the paper reports a study using the authors' Test of Functional Health Literacy in Adults (TOFHLA) to measure the health literacy of 2659 patients in Atlanta (979 patients) and Los Angeles County (1680).

The TOFHLA is a 22-minute test consisting of two parts: a reading comprehension section (50 items) and a numeracy section (17 items). The 17-item numeracy results are scaled up to 50 points possible, so that the two parts are weighted equally.

The reading comprehension part uses a modified Cloze procedure (every fifth to seventh word is omitted; multiple-choice fill-in-the-blank options are provided)[10]. It measures the patient's ability "to read and understand prose passages selected from" instructions for preparation for an upper gastrointestinal tract radiograph series (Gunning Fog grade 4.3[11]), the patient "Rights and Responsibilities" section of a Medicaid

---

[10] From http://olc.spsd.sk.ca/DE/PD/instr/strats/cloze/index.html (accessed 12/29/08): "Instructional Strategies Online" – "Cloze procedure is a technique in which words are deleted from a passage according to a word-count formula or various other criteria. The passage is presented to students, who insert words as they read to complete and construct meaning from the text. This procedure can be used as a diagnostic reading assessment technique." Examples and links are on Web site. Williams et al. (1995) cite the initial paper: Taylor (1953).

[11] From http://www.readabilityformulas.com/gunning-fog-readability-formula.php (accessed 12/29/08) "The Gunning's Fog (or FOG) Readability Formula"—"The ideal score for readability with the Fog index is 7 or 8. Anything above 12 is too hard for most people to read. For instance, *The Bible*, Shakespeare and Mark Twain have Fog Indexes of around 6. The leading magazines, like *Time*, *Newsweek*, and the *Wall Street Journal* average around 11." The Gunning Fog score is calculated as 0.4 times the sum of ASL and PHW, where ASL (average sentence length) is number of words per sentence, and PHW (percentage hard words) is 100 times the number of words with three or more syllables (not counting proper nouns, hyphenated easy words, and two-syllable verbs with an additional "es" or "ed") divided by the total number of words in the passage. The original source is Gunning (1968).

application (grade 10.4), and a standard hospital informed consent document (grade 19.5). The numeracy section tests "the ability of the patients to comprehend" labeled prescription vials, blood glucose test results, clinic appointment slips, and financial information forms. The paper provides results for characterizations of selected numeracy questions (Table 4) but does not include the items themselves.

**Table 4: Results on TOFHLA numeracy questions
(calculated from data in Table 2 of Williams et al. 1995)**

Subject of question	% correct $n = 2659$
How to take medication four times a day	87
How to take medication on an empty stomach	58
How many pills of a prescription should be taken	67
How may times a prescription can be refilled	77
When next appointment is scheduled	64
How to determine financial eligibility	51

A full description of reliability and validity of the TOFHHLA is in a companion paper by Parker et al. (1995)[12]—for internal-consistency reliability, Cronbach alpha was 0.98; for test-retest reliability, Spearman-Brown equal-length coefficient was 0.92; for validity, Spearman rank correlation with the REALM[13] and WRAT-R[14] were 0.84 and 0.74 respectively ($p < 0.001$).

The TOFHLA is copyrighted and is now available from the Center for the Study of Adult Literacy.[15]

Three-Item Numeracy Test – Schwartz et al. (1997)

The goal of the study by Schwartz et al. (1997) was to explore how numeracy affects women's perception of the benefits of mammography when they are given quantitative information. The context was that a National Institutes of Health panel had earlier

---

[12] This paper is in *WoS*, but it was not found in our search. Title and abstract do not use "numeracy".

[13] http://www.hsph.harvard.edu/healthliteracy/doakAB.pdf (accessed 12/29/08): "The Rapid Estimate of Adult Literacy in Medicine (REALM) is a screening instrument to assess an adult patient's ability to read common medical words and lay terms for body parts and illnesses. It is designed to assist medical professionals in estimating a patient's literacy level so that the appropriate level of patient education materials or oral instructions may be used." The standard references are Davis et al. 1991 and 1993.

[14] http://www.cps.nova.edu/~cpphelp/WRAT-R.html (accessed 12/29/08): "The Wide Range Achievement Test−Revised (WRAT-R) is the sixth edition of the popular test that was first published in 1936. Like the earlier versions, the WRAT-R contains three subtests: Reading (recognizing and naming letters and words), Spelling (writing symbols, name, and words), and Arithmetic (solving oral problems and written computations). The authors of the WRAT-R stress that the test is designed to measure basic school codes rather than comprehension, reasoning, and judgment processes." The standard reference is Wilkinson 1993.

[15] http://education.gsu.edu/csal/TOFHLA.htm (accessed 12/29/08).

declined to make a recommendation about screening mammography for women 40−49 years old, deciding instead to advocate that women make their own decisions based on their personal evaluation of risks and benefits. Schwartz et al. (1997) noted that

> This strategy, however, is based on the assumption that patients understand quantitative information…. It is likely that quantitative information is only meaningful to the extent that patients have some facility with basic probability and numerical concepts, a construct called numeracy.

The Schwartz et al. (1997) study consisted of a randomized, cross-sectional survey of 500 female veterans drawn from a New England registry. The women were sent a four-item questionnaire. The first three items were an assessment of numeracy. The last item was an assessment of how the women perceived the benefit of mammography given information about the associated risk reduction. There were four variations in this last question and thus four versions of the questionnaire. The women were each sent one of the four versions. The study investigated the relation between numeracy (assessed by the first three questions) and the ability to use information on risk reduction (assessed by the last question). From the original random sample of 500 women, Schwartz et al. obtained a study sample of 287 usable questionnaires.

Table 5. Three numeracy items from Schwartz et al. (1997)

Question	Answer	% correct $n = 287$
1. Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips? _____ times out of 1,000.	500	54
2 In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS? _____ person(s) out of 1,000.	10	54
3 In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car? _____%.	0.1	20

The three numeracy items in the questionnaire, together with the answers and percentage of the 287 responses that were correct are in Table 5. Only 16% of the participants answered all three questions correctly, and 30% had no correct answers.

The four risk-reduction items each consisted of two parts: a presentation of risk-reduction data (four variations in the way the data were framed) and a two-part question that was the same for all four variations. The two-part question is shown in the first part of Table 6. The four frames with the corresponding correct answers

and results are shown in the lower part of Table 6.

Table 6. Risk-reduction item from Schwartz et al. (1997)

The two-part question

Imagine 1000 women just like you:
  A.   How many will die from breast cancer without mammography?
      _____ out of 1000
  B. How many will die from breast cancer with mammography?
      _____ out of 1000

The four frames –

Data-reduction information	Answer	% correct $n = 287$
• 33% risk reduction from 12 in 1000,(relative risk reduction; baseline risk provided)	B must be less than A by 30-40%	17
• 33% risk reduction (relative risk reduction; no baseline risk provided)		10
• 4 in 1000 risk reduction from 12 in 1000 (absolute risk reduction; baseline risk provided).	A – B must be 4	33
• 4 in 1000 risk reduction (absolute risk reduction; no baseline risk provided)		7

As shown in the last column, most of the women did not apply the risk-reduction information correctly when they were asked to estimate their risk for dying from breast cancer with and without mammography. Accuracy on the risk-reduction question, however, was strongly related to numeracy. For example, 40% of the 45 respondents who answered all three numeracy questions correctly also answered the fourth question correctly (Table 5, column 3). Noting that women with higher numeracy scores tended to be younger and more highly educated, Schwartz et al. used multiple logistic regression analyses to isolate numeracy from age, income, level of education, and framing. The adjustment had little effect (Table 7, columns 4 and 5).

Table 7. Comparison of performance on first three vs. fourth question on the Schwartz et al. (1997) questionnaire.

Numeracy score	n	Accuracy on Question 4 (%)	Odds ratio relative to numeracy scale of 0	
			Without adjustment	Adjusting for age, income, etc
3	45	40	10.9	13.1
2	76	23.7	5.1	7.1
1	79	8.9	1.6	1.3
0	87	5.8		

The three numeracy questions on the Schwartz et al. (1997) test have been incorporated into some other numeracy assessments (e.g., Lipkus et al. 2001, Estrada et al. 2004). According to Nelson et al. (2008, p. 264) "The instrument has moderate internal consistency, with Cronbach's alpha scores ranging from 0.57 to 0.63 in three separate studies (Lipkus et al. 2001)."

The research team of Schwartz et al. applied their three-item numeracy test in a study of the validity of utility assessment (Woloshin et al. 2001)[16]—Given that the standard gamble (SG) and the time trade-off (TTO) utility assessment techniques are inherently quantitative, are utility assessments compromised if test subjects have low numeracy? To find out, Woloshin et al. assessed both the numeracy and utility of current health for 96 women volunteers. To measure utility, they used the SG and TTO and a third, less-numerical approach—a visual analog scale on which respondents wrote a "C" at a position corresponding to how they felt about their current health. The researchers also interviewed the subjects, asking them to rate their current health status. The self reports and the utility for current health showed little correlation for the SG and TTO, and there was a striking effect of numeracy (Table 8). One would expect that higher health ratings would be associated with higher utilities, and thus the r_S-values in Table 6 ought to be positive—and they were for the visual analog utility assessment. For the subjects with low numeracy (0–1 correct answers on the three-item numeracy test), however, the correlations were even in the wrong direction. Only for the most-numerate 38% of the subjects (all three items correct) was the correlation significant for any of the quantitatively demanding measures.

Table 8. Correlation of health rating and utility assessments stratified by numeracy (from Table 2 of Woloshin et al. 2001)

Correlation with health rating	All ($n = 96$)	Number of correct numeracy answers		
		0-1 ($n = 25$)	2 ($n = 35$)	3 ($n = 36$)
SG	0.07	−0.16	+0.10	+0.22
TTO	0.19	-0.13	+0.22	+0.50
Visual	+0.61	+0.82	+0.50	+0.60
Note: Numbers give Spearman correlation r_S. For shaded cells, $P \leq 0.003$; for others, $P \geq 0.2$				

---

[16] Utility quantifies, on a scale of 0 (death) to 1 ("perfect health"), how you value your current health or the possible outcome of a treatment. In the SG assessment, you are asked to imagine living in your current health for the next 10 years vs. trying a painless treatment that will guarantee you perfect health but has some chance of death; a utility of 0.9 means that you chose a gamble that has a 10% chance of death and a 90% chance of perfect health. The TTO assessment asks you to imagine you will spend the remainder of your life in your current state of health or a shorter time in perfect health; a utility of 0.9 means that a person with a life expectancy of 10 years is indifferent between living for 10 years in current health vs. living for 9 years in perfect health. (Woloshin et al. 2001, p. 381)

The three items of the Schwartz et al. (1997) assessment have been used in numerous other research studies in our bibliography (e.g., Sheridan et al. 2003, S. R. Schwartz et al. 2004, Peters et al. 2006; van Tol-Geerdink 2006, Donnelle et al. 2007, Ginde et al., 2008, see appendix). Table 9 attempts to collect results.

Table 9. Results on the Three-Item Numeracy Test of Schwartz (1997)

A. By question (% correct)	(1) $n = 287$	(2) $n = 96$	(3) $n = 463$	(4) $n = 143$	(5) $n = 100$	(6) $n=\sim100$
Likely number of heads (or the dice variation)	54	85	56	72	61	69
Conversion of 1% to 10 in 1000	54	82	60	43	69	35
Conversion of 1 in 1,000 to 0.1%	20	41	21	8	46	15

References: (1) Schwartz et al. 1997; (2) Woloshin et al. 2001; (3) Lipkus et al. 2001; (4) Estrada et al. 2004; (5) Peters et al. 2006; (6) Ginde et al. 2008.

B. By number of correct answers (% of subjects)								
Number correct	(1) $n = 287$	(2) $n = 96$	(3) $n = 463$	(4) $n = 178$	(5) $n = 357$	(6) $n = 18$	(7) $n = 140$	(8) $n = 140$
3	16	38	18	38	2	50	59	16
2	26	36	24	40	28			29
1	28	26	34	15	30	50	41	55
0	30		24	7	41			

References: (1) Schwartz et al. 1997; (2) Woloshin et al. 2001; (3) Lipkus et al. 2001; (4) Schwartz et al. 2005: (5) Sheridan et al. 2003; (6) S. Schwartz et al. 2004; (7) van Tol-Geerdink 2006; (8) Donnelle et al. 2007.

S-TOFHLA – The Short TOFHLA (Baker et al. 1999)

Near the end of their paper on the TOFHLA, Williams et al. (1995) noted the need for a screening tool so that clinicians can identify patients with communication barriers resulting from inadequate literacy skills. They also noted that "the length of the TOFHLA limits its use as a screening tool." Thus they (Baker et al. 1999) created the S-TOFHLA, a shortened version of the TOFHLA. It is a 12-minute (rather than 22-minute) test. The reading comprehension section consists of 36 items involving two prose passages (reduced from 50 items and three passages). The numeracy section consists of four items (down from 17). It tests one's ability to read and understand numerical information as it occurs on, for example, prescription bottles and appointment slips. Subjects are given cue cards or bottles to read (e.g., instructions for taking a medication); then they are asked questions orally. The numeracy items have an average Gunning Fog readability grade of 9.4.

The numeracy items in S-TOFHLA (Table 10) were chosen from the original TOFHLA "based upon the perceived importance and frequency of the task in the health care setting, the proportion of patients who answered items incorrectly, and the perceived

ease of administration" (Baker et al. 1999, p. 35). The developers initially chose five items, but discarded one of them after testing showed poor correlation with the others.

Table 10. The four numeracy items in the S-TOFHLA (compiled from appendix and data in Baker et al. 1999)		
Question	Answer	% correct $n = 211$
A label on a prescription bottle: *Take one tablet by mouth every 6 hours as needed.* Oral question: *If you take your first tablet at 7:00 a. m., when should you take the next one?*	"1:00 pm"	56
A prompt card: *Normal blood sugar is 60–150. Your blood sugar today is 160.* Oral question: *If this was your score, would your blood sugar be normal today?*	"No"	56
An appointment slip with title CLINIC APPOINTMENT and following information: • *CLINIC: Diabetic.* • *LOCATION: 3rd Floor.* • *DAY: Thurs.* • *DATE: April 2nd.* • *HOUR: 10:20.* Oral question: *When is your next appointment?*	"April 2nd" or "Thursday, April 2nd."	59
A label on a prescription bottle: *Take medication on empty stomach one hour before or two to three hours after a meal unless otherwise directed by your doctor.* Oral question: *If you eat lunch at 12:00 noon, and you want to take this medicine before lunch, what time should you take it?*	"11:00" or "Before 11:00."	53

The testing of the STOFHLA was on a convenience sample of 211 patients at an urgent care center and medical clinic at a public hospital in Atlanta. The distribution of correct answers was: four correct, 25%; three correct 23%, two or fewer correct, 52%.

Much of the S-TOFHLA paper (Baker et al. 1999) consists of details on the reliability and validity of the test. As summarized by Nelson et al. (2008):

> The S-TOFHLA haws good internal consistency: Cronbach's alphas for the numeracy items, prose items, and all items combined are 0.68, 0.97, and 0.98, respectively. Like the TOFHLA, the S-TOFHLA is highly correlated with the REALM (0.80). Its reliability and validity are similar to the reliability and validity of the TOFHLA.

The S-TOFHLA has been used as a standard to assess reading comprehension in other studies that developed and applied specific purpose numeracy tests (see Apter et al. 2006, Gong et al. 2007, below).

Expanded, Ten-Item Numeracy Scale (Lipkus et al., 2001)

Lipkus et al. (2001) noted that 36% of the participants in the Schwartz et al. (1997) study and 77% of the participants in the Woloshin et al. (2001) study, which used the Schwartz

et al. (1997) questionnaire, had at least some college. They also noted that only 16% of the participants in the Schwartz et al. (1997) study and 38% of the participants in the Woloshin et al. (2001) study answered all three questions correctly (our Table 9B). The better performance on the latter study may reflect the higher education level, they said, but even so "it is disconcerting that among such a well-educated sample, performance on simple numeracy questions was poor" (Lipkus et al. 2001, p. 38). They were also interested in examining performance on questions phrased in terms of health risks rather than general numeracy.

The paper by Lipkus et al. reports three studies involving a total of 463 "highly educated participants." The participants, men and women aged 40 and older, were recruited via newspaper advertisements to participate in studies pertaining to breast and colon cancer screening. Of the 463 participants, 407 (88%) had at least some college.

Lipkus et al. tested the participants on a two-part, ten-item test, which they called an "expanded numeracy scale"—meaning that it was expanded from the three-item Schwartz et al. (1997) test. The first part of the test—the "general numeracy scale items"—consisted of the same three items as the Schwartz et al. (1997) test except that the coin in the first question was replaced by a six-sided die. For results and how they compare to those of the Schwartz et al. (1997) and Woloshin et al. (2001) studies, see Table 9.

The second part of the test, which Lipkus et al. called the "expanded numeracy scale items," consisted of seven questions[17] written specifically in the context of health risks. The questions and results are shown in the Table 11.

The study included a factor analysis on the 11 questions. The results led the authors to make the following comment (p. 41):

> All the numeracy items were tapping the central construct, global numeracy. Therefore, there was no evidence to suggest that performing mathematical operations in the context of health risks differs from other simple mathematical processes in other contexts. Thus, previously used measures of numeracy probably can be used to assess mathematical performance in the health risk domain. This does not necessarily translate to knowing what a risk means.

Regarding the overall test, Nelson et al. (2008, p. 264) say the following:

> It has adequate internal consistency, with Cronbach's alpha scores ranging from 0.70 to 0.75 in three separate studies (Lipkus et al. 2001). There are no test-retest data available for the scale. Although the numeracy scale is relatively short, it may require up to 30 minutes to complete, making it a frustrating task for some people and calling into question its feasibility in a research or clinical context.

Lipkus et al. concluded that the results of their study "suggest that even highly educated participants have difficulty with relatively simple numeracy questions" (p. 37)—specifically simple probability and questions involving percentages and proportions (p. 41).

---

[17] Some later authors using the Lipkus et al. (2001) scale refer to it as an 11-item test because one of the seven new questions has two parts.

Table 11. The Lipkus et al. (2001) expanded numeracy scale (adapted from Tables 2 and 3 of Lipkus et al. 2001 and Table 1 of Peters et al. 2006)

A. Results by Question			
Question	Answer	% correct	
		Lipkus et al. 2001 $n = 463$	Peters et al. 2006 $n = 100$
1. Which of the following numbers represents the biggest risk of getting a disease? (a) 1 in 100, (b) 1 in 1000, (c) 1 in 10	(c)	78	96
2. Which of the following numbers represents the biggest risk of getting a disease? (a) 1%, (b) 10%. (c) 5%	(b)	84	94
3. If Person's A risk of getting a disease is 1% in ten years, and person B's risk is double that of A's, what is B's risk?	2%	91	83
4. If Person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A's, what is B's risk?	2 out of 100	87	74
5. If the chance of getting a disease is 10%, how many people would be expected to get the disease? • Out of 100? • Out of 1000?	10 100	81 78	90 84
6. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease	20	70	84
7. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?	5 people	49	56

B. Results by number correct (Lipkus et al. 2001)					
Correct	% of 463	Correct	% of 463	Correct	% of 463
8	32	5	12	2	4
7	24	4	8	1	2
6	13	3	5	0	1

The Lipkus et al. (2001) assessment instrument has been used in a variety of studies reported in other papers in the our numeracy bibliography (e.g., Fagerlin et al. 2007, below; Peters et al. 2006, Hibbard et al. 2007, Donnelle et al. 2007, see appendix). In particular Peters et al. (2006) include data on how the 100 participants in their study performed on the questions (Tables 9A and 11). Those participants were recruited using a campus newspaper.

Six-Item, Expanded, Special-Purpose Test (Estrada et al. 2004)

Estrada et al. (2004) hypothesized that patients with low literacy or numeracy skills may have difficulties taking warfarin (Coumadin), which is used to reduce the risk of stroke. They wrote (p. 88),

> Warfarin therapy requires frequent monitoring, dose adjustment and an ability to follow instructions very closely. During follow-up, patients are informed about their international normalized ratio (INR) results and the warfarin dose is adjusted as necessary. Patients are often instructed to cut tablets in half, to take different daily doses, or to use different tablet strengths. Lack of understanding of warfarin dosing increases the risk of an elevated INR by 8-fold and may lead to bleeding complications.

The study involved 143 participants, a prospective cohort that was recruited from two anticoagulation management units (one based at a university, the other at a VA hospital). Literacy and numeracy were independent variables. The main dependent variables were INR variability and the optimal intensity of anticoagulation, both determined from INR data collected during a three-month follow-up period. INR variability was measured as the deviation in the patient's INR from his or her therapeutic range over time. Optimal intensity of anticoagulation was measured with the time-in-range method, which estimates the amount of time a patient's INR is within his or her therapeutic range. With respect to the first dependent variable, "A wider variability of the INR indicates poorer anticoagulation and is one of the strongest predictors of bleeding risks" (p. 89). Regarding the second: "Patients who remain within their therapeutic range are believed to receive optimal anticoagulation management" (p. 89).

**Table 12. Three anticoagulation-control numeracy questions
(adapted from Table 1 of Estrada et al. 2004)**

Question	Answer	% correct $n = 143$
1. If you have 5-mg pills of Coumadin and you take 7.5 mg a day, how many of those pills should you take every day? _____ pills	1.5	62
2. You have 5-mg pills of Coumadin and you take 7.5 mg a day. If you have 9 pills left, would you have enough for one week? Yes/No	No	52
3. Your normal INR should be 2 to 3. If your INR today is 3.5, would your INR be: Low/Normal/High	High	64

To measure numeracy, Estrada et al. used a six-item test.[18] Like Lipkus et al. (2001), they started with the three "general numeracy questions" from Schwartz et al. (1997) and added new questions for their specific study (Table 12). The results for the three general-numeracy questions from Schwartz et al. (1997) are included in Table 9B.

---

[18] The paper does not contain information about reliability or construct validity.

For the overall six-item expanded test, the distribution of correct answers was: 5–6 correct, 8%; 3–4 correct, 34%; 1–2 correct, 35%; and none correct, 13%.

Estrada et al. examined the relation between the level of literacy and numeracy and the anticoagulation-management indicators (INR variability and time in range) with Spearman rank test, using multiple linear regression to adjust for age. For numeracy and INR variability, they found (p. 90-91):

> … the INR variability was higher among patients with lower numeracy skills ($r_S = -0.24$, $P = 0.004$).… For example, the INR variability was 46% higher in patients at the lower numeracy level as compared with patients at the highest numeracy level. The relationship between INR variability and numeracy skills were similar for the [three items of Table 3] ($r_S = -0.20$, $P = 0.02$) and the [three items of Table 8] ($r_S = -0.22$, $P = 0.009$). After adjusting for age … the INR variability was higher among patients with lower-level literacy ($P = 0.06$) or numeracy skills ($P = 0.03$).

For numeracy and time in range, they found (p. 91):

> Patients at lower numeracy levels had a trend toward less time spent in range ($r_S = 0.14$, $P = 0.10$). For example, patients who could not answer any of the numeracy questions spent less time in range (55.8%) as compared with patients who answered 5 or 6 questions correctly (67.8%). The relationship between time in range and numeracy skills were similar for the [three items of Table 3] ($r_S = 0.06$, $P = 0.45$) and the [three items of Table 8] ($r_S = 0.09$, $P = 0.26$). After adjusting for age … the time in range was similar among patients at different literacy ($P = 0.71$) or numeracy ($P = 0.35$).

The authors concluded (p. 92) that low numeracy is prevalent and associated with poor anticoagulation control. "Literacy and numeracy may be the primary factors responsible for the strength of the association or they could be confounders for other factors, including age."

MDIT – Medical Data Interpretation Test (Schwartz et al., 2005)

The purpose of the work reported by Schwartz et al. (2005) was to develop and establish the validity of an assessment tool to measure medical data interpretation skills. They designed the MDIT to be (p. 295)

> … a comprehensive test of skills needed to make sense of the kind of real-world health information people routinely encounter in direct-to-consumer prescription drug advertisements, news media reports, and statements physicians typically use in conversation with patients about medical risks.

The authors clearly mean for the MDIT to be a step beyond the tools of earlier studies such as theirs (Schwartz et al. 1997) and Lipkus et al. (2002), which assess patients' ability to perform simple mathematical operations such as converting between percentages and proportions. They stress that the new measure (p. 291) "… examines the ability to compare risks and put risk estimates into context (i.e., to see how specific data fit into broader health concerns and to know what additional information is necessary to give a medical statistic meaning)."

Table 13. Example single-question assessment items from the MDIT (compiled from Table 2 and appendix in Schwartz et al. 2005)

Category / purpose	Question	Answer	% correct $n = 178$
Knowledge basis for comparison / Know that a denominator is needed to calculate risk	*About 51,000 Americans will be diagnosed with melanoma (the most serious skin cancer) this year.* What is your best guess about an American's chance of being diagnosed with melanoma in the next year? (a) 51,000, (b) 51,000 divided by the number of Americans, (c) Don't know how to figure this out	(b)	75
Knowledge basis for comparison / Know that denominators are needed to compare risks in two groups).	*A study finds that there were 30 deaths among people who eat broccoli regularly compared to 100 deaths among people who don't eat broccoli at all.* According to this study, which statement best describes how eating broccoli relates to death? (a) Lowers risk of death. (b) Doesn't change risk of death. (c) Raises the risk of death. (d) Can't tell from this information.	(d)	45
Comparison task / Select "1 in 206" as larger risk than "1 in 407"	*Mrs. Smith is told she has a 1 in 296 chance of dying from cancer and a 1 in 407 chance of dying from a stroke.* Which is bigger, Mrs. Smith's chance of dying from a stroke or cancer? (a) Stroke. (b) Cancer. (c) Chances are the same	(b)	85
Calculation related to comparison / calculate relative risk reduction from two absolute risks	*In a new study, people either took pill X or placebo (a sugar pill). 3% of people taking placebo died; 1% of people taking pill X died.* Which statement is correct about how pill X changes the chance of death? (a) Lowers by 66%. (b) Lowers by 33%. (c) Raises by 33%. (d) Raises by 66%.	(a)	52
Context for comparisons / Know that age of individuals in the source data is needed):	*Your doctor says there is a 10% risk of dying of pneumonia.* Which information best helps you understand how this risk *applies to you*? (a) Most people who die from pneumonia are 75 years or older. (b) More than 110,000 people get pneumonia each year. (c) Pneumonia is one of the most common reasons for hospitalization. (d) About 15,000 people die from pneumonia each year.	(a)	60

The MDIT consists of 18 items involving 20 questions. It was evaluated on a sample of 178 people recruited from advertisements in local newspapers, an outpatient clinic, and an open house at a hospital. Test results on the 18-item MDIT were scaled 0

to 100, which gave an average of 61 and standard deviation of 17. Test-retest reliability was established by correlation of scores for a subsample of 84 people who retook the test after two weeks (Pearson $r = 0.67$; target was $r > 0.6$). Internal-consistency reliability (degree to which items measure a single construct) was established with a Cronbach's alpha of 0.71 (goal was $0.7-0.8$, a moderate value because "we sought items that measured the domain of data interpretation but included a wide range of skills," Schwartz et al. 2005, p. 292). Content validity was formally assessed by 15 Dartmouth Medical School faculty who teach evidence-based medicine. Construct validity was assessed by testing whether MDIT scores discriminate among groups with different abilities, one of which was numeracy as measured by the three-item scale of Schwartz et al. (1997). The results on the three-item test are included in Table 9B. People with highest numeracy (all three answers correct) scored higher on the MDIT than people with lowest numeracy (no answers correct)—71 vs. 36, $P < 0.001$ (Schwartz et al. 2005, Table 3).

The paper (Schwartz et al. 2005, Table 2) breaks the 18 items down into 4 categories and gives the purpose of each question. The four categories are: knowledge basis for comparisons (5 items), comparison tasks (4), calculations related to comparisons (5), and context for comparisons (4). Example assessment items with the purpose and category are shown in Table 13.

In addition to single-question assessment items such as the examples in Table 13, three of the four MDIT items in the comparison-tasks category consist of paired questions that are assessed on the internal consistency of the answers. For example, two items are made from the three questions in Table 14.

Table 14. Example paired-question assessment items from the MDIT (compiled from Table 2 and appendix in Schwartz et al. 2005)

Category / purpose	Question Premise: *Mr. Newman is a healthy 40-year-old man who does not smoke.*	Answer	% correct $n = 178$
	What is your best guess about Mr. Newman's chance of dying from a heart attack in 10 years? (a) 1 in 1,000; (b) 5 in 1,000; (c) 30 in 1,000; (d) 80 in 1,000; and (e) 250 in 1,000.		
Comparison task / Select a larger risk estimate for deaths from all causes than death from a specific cause.	What is your best guess about Mr. Newman's chance of dying for any reason in the next 10 years? (a) 1 in 1,000; (b) 5 in 1,000; (c) 30 in 1,000; (d) 80 in 1,000; and (e) 250 in 1,000.	Must be larger than previous answer	30
Comparison task / select a larger risk estimate for a 20-year than 10-year risk"	What is your best guess about his chance of dying for any reason in the next 20 years? (a) 1 in 1,000; (b) 5 in 1,000; (c) 30 in 1,000; (d) 80 in 1,000; and (e) 250 in 1,000.	Must be larger than previous answer	39

Most of the respondents gave the same answer for the paired questions; i.e., 62% estimated the all-cause risk to be the same as the disease-specific risk, and 56% estimated the 20-year risk to be the same as the 10-year risk.

The full MDIT is a 5-page appendix in Schwartz et al. (2005). It is separately available online[19] both as a Word document and a pdf. There is a validated Dutch translation (Smerecnik and Mesters 2007).

ANQ – Asthma Numeracy Questionnaire (Apter et al. 2006)

A four-item asthma numeracy questionnaire (Apter et al. 2006) was developed by specialist clinicians from common recommendations made to patients with moderate or severe asthma. The aim was to develop and apply an instrument to test patients' abilities to interpret specific self-management instructions and to provide information that would guide clinicians to adjust instructions to better communicate with individual patients.

Table 15. Numeracy questions from the Asthma Numeracy Test (adapted from Table 2 of Apter et al. 2006)

Here are some examples of statements or questions that patients might hear in a doctor's office:

Question	Answer	% correct $n = 73$
Your doctor asks you to take 30 mg of prednisone every day for a week. The pharmacist gives you a bottle of 5-mg tablets. How many pills should you take each day?	6	61
If a patient has a 1% chance of developing osteoporosis or bone loss, that means: a. Out of 1,000 patients, one will develop bone loss. b. Out of 100 patients, one will develop bone loss. c. Out of 10 patients, one will develop bone loss. d. Out of 5 patients, one will develop bone loss. e. The patient will develop bone loss. f. The patient will never develop bone loss.	(b)	28
You have a peak flow meter. Your Danger or Red Zone is 50% of your best reading. Your best reading is 400 L/min. What is your Danger Zone? _____L/min or less.	200	52
You are told the Green Zone (the OK zone) is a reading between 80% and 100% of your best reading. Your Worry Zone is between 50% and 80% of your best reading. Your best reading is 400 L/min. When are your readings in the Worry Zone? a. Between 300 and 400 L/min. b. Between 200 and 320 L/min. c. Between 200 and 300 L/min. d. Between 240 and 320 L/min. e. Between 100 and 300 L/min.	(b)	21

The authors used proportional odds ordinal logistic regression analyses to relate ANQ scores to asthma-severity variables. The asthma-severity variables included lifetime hospitalizations and visits to emergency departments (ED) using four categories

---

[19] http://129.170.61.41/research_tools.php (accessed 12/29/08).

(0, 1, 2, and 3 or more). They controlled for age, sex, educational attainment and household income. They found negative and significant associations between the asthma-severity variables and the numeracy scores: estimated odds ratio, 0.546 ($P = 0.012$) for hospitalizations and 0.563 ($P = 0.004$) for ED visits. In contrast, there was not a significant association between the S-TOFHLA scores and the asthma-severity measures (P values between 0.228 and 0.783). According to these researchers, the study "(demonstrates) that risk assessment and percentages directly tied to immediately relevant self-management skills are important aspects of medical instructions that may not be well understood by patients" (p. 709). Moreover, "Widely used measures of health literacy do not identify many adults who lack numerical skills needed for asthma self-management" (p. 710).

SNS – The Subjective Numeracy Scale (Fagerlin et al., 2007)

Noting that research participants may be averse to objective numeracy tests such as the numeracy part of the TOFHLA and the numeracy scales of Schwartz et al. (1997) and Lipkus et al. (2001), and that such tests would not be easily administered in phone or Internet surveys, Fagerlin et al. (2007) developed a subjective scale. The test was subjective in that the questions asked the respondents to assess their ability to perform various mathematical tasks and for their preferences regarding data presentation.

Development of the test started with a focus group and went through several rounds of testing. In the process, the researchers tested 49 subjective numeracy questions against objective numeracy items from Lipkus et al. (2001). They selected eight questions for the final SNS test. The first four questions ask respondents about their ability to perform various mathematical tasks. The second four questions ask for their preferences regarding the presentation of numerical and probabilistic information.

The eight questions are shown in Table 16 with results (mean and standard deviation) from 287 participants recruited from a VA hospital waiting room. Overall, participants' scores ranged from 1.00 to 6.00, with mean and standard deviation of 4.03 and 1.04 respectively, and a slight negative skewness (−0.36). Cronbach's alpha for internal-consistency reliability was 0.82. No information was given about test-retest reliability.

Pearson bivariate correlations were conducted between each of the SNS items and the Lipkus et al. (2001) objective numeracy scale (Table 16, column 3). There was a significant correlation between the overall subjective and objective scales ($r = 0.53$, $P < 0.01$; after correcting for the reliability of the two scales, $r = 0.68$). There was also significant correlation of the objective scale and each part of the SNS—the first four, self-assessment ability questions ($r = 0.47$, $P < 0.01$; with correction, $r = 0.60$) and the second four, preference questions ($r = 0.44$, $P < 0.01$; with correction, $r = 0.56$).

In a separate study of preference for the subjective vs. the objective tests, 52 people were recruited from the cafeteria of a university hospital. They were randomly selected to take either the 11-item Lipkus et al. (2001) assessment or the new 8-item SNS. They were asked to rate their reactions to their survey by a 6-point Likert scale in four areas: enjoyment, annoyance, stress and frustration. The SNS was preferred in all four areas.

There was significantly less annoyance (2.42 vs. 3.42) and less stress (1.62 vs. 2.69) with the subjective test ($P < 0.01$ for the *t*-statistic in each case). There was also less frustration (1.92 vs. 2.88, $P < 0.05$) and more enjoyment (4.27 vs. 3.96, $P > 0.10$) with the subjective test. The participants completed the SNS more quickly than the objective test (5.03 vs. 7.49 minutes, $P < 0.01$).

Table 16. The eight questions of the Subjective Numeracy Scale (adapted from Table 2 of Fagerlin et al. 2007)

Question	Result	Correlation with objective scale
How good are you at working with fractions? (1 = *not at all good*, 6 = *extremely good*)	3.67 ± 1.51	0.41
How good are you at working with percentages? (1 = *not at all good*, 6 = *extremely good*)	3.92 ± 1.47	0.43
How good are you at calculating a 15% tip? (1 = *not at all good*, 6 = *extremely good*)	4.20 ± 1.54	0.41
How good are you at figuring out how much a shirt will cost if it is 25% off? (1 = *not at all good*, 6 = *extremely good*)	4.58 ± 1.40	0.34
When reading a newspaper, how helpful do you find tables and graphs that are parts of a story? (1 = *not at all*, 6 = *extremely*)	3.83 ± 1.43	0.26
When people tell you the chance of something happening, do you prefer that they use *words* ("it rarely happens") or *numbers* ("there's a 1% chance")? (1 = *always prefer words*, 6 = *always prefer numbers*)	3.53 ± 1.82	0.33
When you hear a weather forecast, do you prefer predictions using *percentages* (e.g., "there will be a 20% chance of rain today") or predictions using only *words* (e.g., "there is a small chance of rain today")? (1 = *always prefer percentages*, 6 = *always prefer words*; reverse coded)	3.06 ± 1.90	0.27
How often do you find numerical information to be useful? (1 = *never*, 6 = *very often*)	4.05 ± 1.46	0.30

In a companion SNS-validation paper, Zikmund-Fisher et al. (2007) report the results of studies of the association of SNS-assessed numeracy and the ability of participants to comprehend risk communications and meaningfully complete utility elicitations. The studies examined the participants' ability to

- recall risk information presented textually (following up Schwartz et al., 1997), using statistics in the context of angina and angioplasty;
- recall risk information using the same statistics presented as pictographs in which probabilities are shown as 5×20 matrices of small figures;
- comprehend survival curves in the context of comparing the effect of two different pills on a hypothetical disease;

- perform a time-tradeoff (TTO) with internally consistent responses (following up on Woloshin et al., 2001); and
- give consistent answers in a person-tradeoff (PTO) task—patients were asked to imagine that they were the executive director of a health care organization and then a question such as "how many patients need to be cured of mild shortness of breath to bring the same amount of benefit as curing 100 patients of severe shortness of breath" (p. 669).

In some of the studies, participants completed the Lipkus et al. objective numeracy measure as well as the SNS. Some were administered on the Internet after recruitment using an Internet tool (offering a drawing for cash prizes). Logistic regression analysis was used to test the association of numeracy scores to performance on the risk communications and utility elicitations.

Zigmund-Fisher et al. say the following about the comparisons (p. 670):

> Survey participants' responses to the questions on the SNS, a self-assessment of both numerical aptitude and preferences for numerical information, are significantly related to performance on numerical information-processing tasks relevant to medical decision making. Individuals with low numeracy levels (as measured by both the SNS and the objective numeracy measure) are less likely to (1) recall risk information presented textually or in pictographs, (2) comprehend risk information displayed in survival curves, and (3) effectively complete utility elicitation measures.
>
> ….
>
> Although it was neither our intent nor our expectation that the SNS would match objective tests of numerical ability, the SNS did hold up reasonably well compared with an established numeracy measure. Mathematics tests, such as the standard objective numeracy questions, assess one's ability to do a specific task. Our self-perceptions of skill are necessarily one step removed from aptitude. Nevertheless, it seems clear that the SNS significantly predicts the same behaviors as objective numeracy measures do. Just how equivalent the SNS and objective numeracy measures will be in a particular application is, of course, context dependent. Similarly, the minimum SNS score required to predict satisfactory performance is likely to vary from situation to situation. Still, we believe that the SNS can effectively proxy for objective numeracy measures in many circumstances, reducing respondent burden while maintaining significant predictive validity.

Soon after these researchers developed (Fagerlin et al. 2007) and validated (Zigmund-Fisher et al. 2007) the SNS, they incorporated it into studies using Internet-administered surveys. One of their papers (Zigmud-Fisher et al. 2008) explored the issue of how to present quantitative information to patients about the risks of side effects associated with a drug treatment. They reported two experiments comparing the presentation of total risk statistics (risk of each side-effect condition with and without the drug) to presentation of the incremental risk—i.e., a presentation explicitly stating or graphically highlighting the difference between the baseline risk and the with-treatment risk of the side effect. "On a scale of 0 ('not at all worried') to 10 ('extremely worried'), how worried would you be about [side effect] if you took [the drug]?"

The first experiment was a comparison of side-by-side displays of total risk vs. a presentation showing the incremental risk by using a sequence of displays (risk without

drug followed by additional risk with drug). Results showed that the latter presentation significantly lowered participant worry about complications. The experiment raised the question of whether the difference was due to the sequential presentation or the explicitness of the incremental risk.

The second experiment was designed to examine the effect of incremental risk and sequential displays individually. It also included a second outcome measure, the perceived likelihood of experiencing the side effect if one took the drug (on a scale of 0 for 'not at all likely' to 10 for 'extremely likely'). This second experiment found that the effect "on both perceived likelihood and worry derives primarily from the incremental risk framing rather than from sequential presentation" (p. 108). It also showed a numeracy effect[20] (p. 119):

> Unlike in [the first experiment], individual differences in numeracy did consistently influence survey responses in [the second experiment]: participants with higher numeracy scores perceived significantly less risk and were less worried about side effects than participants with lower numeracy scores. We also observed a strong effect of race: In every analysis, African American participants perceived the side effects … to be significantly more likely and more worrisome than did White participants. Neither numeracy nor race, however, showed any consistent interaction with the incremental risk or sequential presentation manipulations.

In another application paper, the research group (Damschroder et al. 2008) investigated whether people would give higher utility (quality of life) ratings for chronic health conditions if they performed a three-question exercise that encouraged them to think through their ability to adapt to difficult situations. The study involved a total of 1,653 Internet participants, half of whom did the adaptation exercise before the utility elicitation. Between-subjects t tests were used to test the differences in utility ratings given by participants who did the adaptation exercise vs. those who did not for each of four conditions (below-the-knee amputation, paraplegia, severe back pain, colostomy) and each of two elicitation methods (TTO, SG). The study produced a null result: "In contrast to earlier research, regardless of elicitation method or health condition, the adaptation exercise did not influence respondents' valuations" (p. 396). Within the details, the authors note a numeracy effect: 19% of the people assigned to the TTO elicitation did not complete it, and 26% of the people assigned to the SG elicitation did not complete it. The authors report that "respondents who did not complete the survey had lower subjective numeracy than those who did, $t(1,540) = -4.8$, $p < 0.001$" (p. 396).

TOFHLiD – Test of Functional Health Literacy in Dentistry (Gong et al. 2007)

The TOFHLiD assessment instrument is aimed at parents of pediatric dental patients. Like the TOFHLA which was used as a template, the TOFHLiD consists of a reading

---

[20] $N = 1,393$. Participants imagine they have had mild epilepsy; hypothetical medication would eliminate seizures but carries risk of three complications: strokes, headaches, and increased frequency of colds. ANOVA. F-statistic for SNS score for both likelihood and worry for each side condition and all side effects range from 13.3 to 18.4, all with $p < 0.001$ (Table 5).

comprehension section followed by a numeracy section. The test was made by a panel of pediatric and public health dentists who reviewed and selected material from patient-education and instructional materials at the university dental clinic. The test was given to caregivers (mostly parents) of children less than 15 years old.

The numeracy section consists of 12 questions: five using instructions for fluoridated toothpaste; three about a pediatric dental clinic appointment; two using bottle prescription labels for fluoride drops; and two for fluoride tablets. The paper includes the five toothpaste questions (Table 17).

Table 17. A set of questions from the TOFHLiD (adapted from Figure 1 of Gong et al. 2007)

Instructions: • Adults and children 2 years and older: brush teeth thoroughly, preferably after each meal or at least twice a day, or as directed by a dentist or doctor.   • Children under 6 years of age: use only a pea-sized amount of toothpaste.   • Instruct children under 6 years in good rinsing habits (to reduce swallowing).   • Supervise children as necessary until capable of using without supervision.   • Children under 2 years of age: ask a dentist or doctor.
Questions: 1. What is the minimum number of times that teeth should be brushed?   2. Should teeth be brushed after breakfast?   3. Should a child who is 1½ years of age use this toothpaste?   4. What amount of toothpaste should be used if a child is less than 6 years old?   5. From a display of five pictures of toothbrushes with different amounts of toothpaste [not shown here], which one matches the amount of toothpaste that should be used for a child younger than 6 years of age?

Gong et al. examine the reliability and validity of the test. The 102 participants were each tested with two instruments for dental health literacy (TOFHLiD and REALD-99[21]) and two instruments for medical health literacy (TOFHLD and REALM). Internal-consistency reliability was determined by Cronbach's alpha. Construct validity was examined by the multitrait-multimethod (MTMM) matrix, which is shown in Table 18. The values in the green cells of the main diagonal are the results for Cronbach's alphas of the respective instruments; Cronbach's alphas for the reading comprehension and numeracy sections (not shown) were 0.65 and 0.59, respectively. The other values in the matrix are the pairwise Spearman correlation coefficients of the four instruments. The values in the olive-colored cells show high correlation of different measures of the same construct (oral health in the case of third row, first column of the matrix); these cells test for convergent validity. The blue and purple cells (same measure/different construct and different measure/different construct, respectively) can give evidence of discriminant validity. As desired, these values relating tests of different constructs (oral and medical health) in the TOFHLiD column were larger than zero and less than 0.82, the value

---

[21] Rapid Estimate of Adult Literacy in Dentistry, a dental health literacy word-recognition instrument: Richman et al. (2007).

relating tests of the same construct. However, the authors were troubled that the two values (0.52 and 0.53) were similar; the good result would have been that the value relating tests that were different in both method and construct (purple cell) would be the smallest. They concluded (p. 112) that the "TOFHLiD demonstrates good convergent validity but only moderate ability to discriminate between dental and health literacy." Moreover, "the value for Cronbach's alpha is at the lower range of values needed to demonstrate acceptable internal consistency."

Table 18. MTMM matrix for TOFHLiD
(adapted from Table 4 of Gong et al. 2007)

		Reading and numeracy		Word recognition	
		TOFHLiD	TOFHLA	REALD-99	REALM
Reading and numeracy	TOFHLiD	0.63			
	TOFHLA	0.52	0.67		
Word recognition	REALD-99	0.82	0.39	0.86	
	REALM	0.53	0.82	0.79	0.92

Gong et al. also assessed predictive validity of the TOFHLiD by comparing three separate outcome variables to TOFHLiD and TOFHLD scores. The oral health outcomes were participants' assessments of their oral health-related quality of life (linear regression), participants' global ratings of their oral health (logistic regression), and participants' global ratings of the child's oral health (logistic regression). They found (p. 111) "that TOFHLA was not associated with any of the three oral health outcomes, while TOFHLiD was associated with one (the participants' assessments of their oral health-related quality of life)."

The conclusions of the paper were those of a progress report. Although oral health literacy and medical health literacy appear to be correlated, the correlation does not appear strong enough that the authors would recommend using a medical health literacy test to assess oral health literacy. But their TOFHLiD is not yet ready for clinical or public health practice; the results "provide a solid foundation for more research" (p. 112). In particular (p. 111), the "TOFHLiD has sufficient discriminant power to justify further exploration."

DNT – Diabetes Numeracy Test (Huizinga et al. 2008; Cavanaugh et al. 2008)

The Diabetes Numeracy Test (DNT) was designed to assess numeracy skills that patients need for their glucose monitoring and management. Quoting Huizinga et al. (2008, p. 2),

> For a patient with diabetes, numeracy is needed to interpret blood glucose meter data, properly administer medications and follow nutritional recommendations. For example, a patient with diabetes may need numeracy to calculate their carbohydrate intake and adjust their insulin based on carbohydrates and/or current blood glucose level. Poor numeracy skills in patients with diabetes could lead to suboptimal glycemic control, increased hypoglycemic episodes or widely varying glucose values.

The paper by Huizinga et al. (2008) describes the development, internal reliability, and construct validity of the DNT. The 43 items of the DNT were culled from 70 developed by a group of experts (including diabetes specialists and educators, primary care providers, registered dietitians, and literacy and numeracy experts). It was tested on a convenience sample of 398 patients recruited from clinic visits. The test took an average of 33 minutes, and the average score was 61%. It was found to have high internal-consistency reliability (Kuder-Richardson 20 coefficient was 0.95) and was judged to have excellent construct validity through consistency with an *a priori* model of expected correlations hypothesized by an expert panel. Spearman rank correlations were: 0.52 with education; 0.51 with income; 0.54 with results on the REALM; 0.62 with results on the WRAT-3R; 0.71 with results on the DKT[22]; and 0.04 with insulin use; *P* was <0.0001 for all hypothesized correlations except the last, for which *P* was 0.43.

The companion paper by Cavanaugh et al. (2008) gives more information on the connection between numeracy (as indicated by the WRAT, which measures calculation skills) and the DNT. Of the 398 participants, 69% (276) measured below 9th grade level on the WRAT. High-numeracy participants ($\geq$ 9th grade level) scored consistently higher on the DNT questions (Table 19).

Table 19. Percent correct on selected DNT tasks stratified by WRAT result (adapted from Table 2 of Cavanaugh et al. 2008)

Sample Tasks	Overall	Numeracy Level	
		<9th grade	$\geq$ 9th grade
Nutrition			
1. Calculate total grams of carbohydrate in one container of snack chips	44	31	75
Exercise			
2. Calculate carbohydrate intake needs for planned exercise duration	64	50	97
Glucose monitoring			
3. Identify values within target range of 60−120 mg/dL	74	67	88
4. Calculate date needed to refill strips	62	50	89
Medications/insulin			
5. Mark 54 units on a 100-unit syringe	66	56	90
6. Calculate insulin needed for carbohydrate intake	65	54	92
7. Titrate of oral hyperglycemic medication			
8. Interpret insulin correction scale table (i.e., sliding scale	65	53	92
	85	78	100
9. Calculate insulin dose, adjusted for blood glucose level and carbohydrate intake	41	28	72
10. Understand titration instructions for long-acting insulin regimen	38	25	68
Median total correct	65	51	83

---

[22]Diabetes Knowledge Test, which consists of a 14-item general test (e.g., nutrition, glucose control) and a 9-item insulin-use subscale. It takes about 15 minutes to complete. Fitzgerald et al. 1998.

Huizinga et al. (2008) include a useful table classifying the content of the DNT. They list five diabetes care domains: nutrition (9 items), exercise (4), blood glucose monitoring (4), oral medication use (5), and insulin use (11). They identify seven numeracy domains: addition (2 items); subtraction (1); multiplication (5); division (6); fractions/decimals (4); multi-step mathematics (13); time (3); numeration/counting/hierarchy (10).

Between them, the two papers (Huizinga et al. 2008; Cavanaugh et al. 2008) include a total of six sample questions from the DNT (Table 20).

Table 20. Six sample questions from the Diabetes Numeracy Test (Huizinga et al. 2008; Cavanaugh et al. 2008)

Question	Answer	% correct
You are told to follow the sliding scale shown in the following table. The sliding scale indicates the amount of insulin you take based upon your blood sugar levels. **If Blood sugar is:** / **Units of Insulin** 130-180 / 0 181-230 / 1 231-280 / 2 281-330 / 3 331-380 / 4 How much insulin should you take for a blood sugar of 295?	3 units	85
If you ate the entire bag of chips, how many total grams of carbohydrates would you eat? **Nutrition Facts** Serving size 1 oz (28 g/About 10 chips) Servings Per Container 3.5 Amount Per Serving Calories 140 / Calories from Fat 60 / % Daily Value Total Fat 6 g / 10% Saturated Fat 0.5 g / 4% Cholesterol 0 mg / 0% Sodium 150 mg / 7% Total Carbohydrate 18 g / 6%	63 grams	44%
Your target blood sugar is between 60 and 120. Please circle the values below that are in the target range (circle all that apply) 55 145 118	118	74
You have a prescription for Metformin 500-mg tablets. The label says, "Take 1 tablet with supper each night for the first week. Then, increase by 1 tablet each week for a total of 4	2 tablets	65

tablets daily with supper." How many tablets should you take with supper each night the second week?		
You check your blood sugar just before eating. You take 1 unit of insulin for every 10 grams of carbohydrates you eat. You are also given the sliding scale shown below. The sliding scale indicates the amount of insulin you should add to your usual dose on the basis of your blood sugar levels: • If your blood sugar is greater than 120 points at breakfast, lunch, or supper, add 2 units of insulin. • If your blood sugar is greater than 150 points at breakfast, lunch, or supper, add 4 units of insulin. • If your blood sugar is greater than 180 points at breakfast, lunch, or supper, add 6 units of insulin.	7 units	33

1 unit of insulin per 10 grams of carbohydrates at meals			
If Blood sugar is:	Breakfast	Lunch	Supper
>120	+2	+2	+2
>150	+4	+4	+4
>180	+6	+6	+6

Your blood sugar is 140 and you will eat 50 grams of carbohydrates at lunch. How much total insulin do you need to take?

The paper by Cavanaugh et al. (2008) examined predictive validity: the association between DNT results and glucose control. The primary measure of a glucose control was the patient's most recent level of hemoglobin A_{1C}. They found it to be higher for participants in the lowest quartile of DNT scores than for those in the highest quartile (7.6% vs. 7.1%; $P = 0.119$). Further, they reported (p. 10), "Regression analysis examining the relationship between DNT score and hemoglobin A_{1C} level, adjusted for age, sex, race, income, diabetes type, years of diabetes diagnosis, and clinic site showed that DNT score was modestly associated with hemoglobin A_{1C} level." A 10-percentage point decrease in the DNT score predicted a rise of 0.09 percentage points in the hemoglobin A_{1C} level ($P = 0.027$). By comparison, they noted, a 5-year increase in age predicted a decrease of 0.17 percentage points in the A_{1C} level ($P < 0.001$).

In the final stage of the study reported by Huizinga et al. (2008), a shortened, 15-item version, the DNT15, was created and verified through split sample analysis. The DNT15, which was developed for more convenient clinical use, performed similar to the full DNT on the internal-consistency and construct validity tests. It was highly correlated with the DNT (Spearman correlations were 0.96 and 0.97 for the development and confirmation samples, respectively). The DNT15 assessments were completed in 10−15 minutes.

Both the full DNT and the DNT-15 are available online.[23]

---

[23] http://www.mc.vanderbilt.edu/diabetes/drtc/preventionandcontrol/tools.php (accessed 12/29/08).

Applications

The papers by Baker et al. (1999), Schwartz et al. (2005), Fagerlin (2007), Gong et al. (2007) and Huizinga et al. (2008) in the preceding compilation are specifically about the development and testing of numeracy tests: the S-TOHFLA, MDIT, SNS, TOHFLiD and DNT, respectively. The first three are general-purpose health-numeracy tests, and the last two are specifically aimed at a particular context: dentistry and diabetes, respectively.

The papers by Williams et al. (1995), Schwartz et al. (1997), Lipkus et al. (2001), Estrada et al. (2004), and Apter et al. (2006) develop and report the assessment instrument and then apply it to a research question. Between them the papers span a wide range of evaluations: the adequacy of functional health literacy amongst the patients in urban hospitals (Williams et al. 1995); the ability of patients to apply risk-reduction information to evaluate the benefit of mammography screening (Schwartz et al. 1997); the ability of patients with some college education to evaluate the benefits of breast and colon cancer screening (Lipkus et al., 2001); the ability of patients to monitor and adjust their medication dosages for anticoagulation control (Estrada et al. 2004); and the ability of patients to interpret instructions to self-manage their asthma (Apter et al. 2006). Additionally the compilation includes some papers in which the members of the team that developed the assessment tool apply it in a separate paper. Those applications include studies of the ability of patients to accurately complete utility assessments (Woloshin et al. 2001) and the ability of diabetes patients to manage their glucose levels (Cavanaugh et al. 2008).

The appendix briefly describes 10 additional papers that extend the range of applications of the various health numeracy assessment tools even further.

- How does numeracy affect patients' decision making (Peters et al. 2006)?
- Does numeracy contribute to how well patients comprehend and compare health-care performance reports (Hibbard et al. 2007)?
- How does one communicate potential benefits and harms of treatments to patients (Sheridan et al. 2003)?
- Does numeracy affect patients' perception of the risk of colon cancer (Kelly et al. 2007)?
- How does numeracy affect the validity of utility assessments for patients with head and neck cancer (Schwartz et al. 2004)?
- Which patients want to be involved in making choices in the treatment of their prostate cancer (Van Tol-Geerdink et al. 2006)?
- How does low numeracy affect incorrect medication dosing of young children (Yin et al. 2007)?
- How prevalent is limited numeracy amongst patients at emergency departments? (Ginde et al. 2008)

- Does numeracy affect how patients (students) rate physicians who use computer programs to arrive at a treatment decision (Arkes et al. 2007)?
- How does numeracy relate to prose literacy, education and math anxiety (Donnelle et al. 2007)?

The collection of articles shows the widespread use of numeracy assessments in research concerning ways of communicating effectively with patients; how patients make medical decisions; how they look at risk; how well they can self-manage their medical conditions; and how well they, as participants in research, understand what the research items are asking. The medical conditions run the gamut: breast, colon, prostate, head/neck cancer; cardiovascular problems; diabetes; asthma.

Other Assessment Instruments

The section on numeracy assessment in the review paper by Nelson et al. (2008) covers all the validated general-purpose instruments we found in our search of *WoS* (i.e., TOHFLA and S-TOHFLA, the Schwartz et al. 1997 and Lipkus 2001 scales, the MDIT, and the SNS). It adds two others we did not find:

- The Newest Vital Sign (Weiss et al. 2005)—six questions about data from a nutrition label on a pint of ice cream.
- The STAT-Confidence scale (Woloshin et al. 2005)—three subjective questions about how one reacts to medical statistics.

Both papers are in the *WoS* database, but they do not use the term "numeracy" in the search areas.

We note that some researchers (e.g., Williams et al. 1995; Rothman et al. 2006; Cavanaugh et al. 2008) used the Wide Range Achievement Test (Wilkinson 1993) as a measure of numeracy. One study (Donnelle et al. 2007) assessed math anxiety with the Abbreviated Math Anxiety Scale (Hopko et. al. 2003).

Assessment Items (EXTRACTED PRODUCT)

Our mining exercise in the WoS recovered 48 assessment items from the health numeracy literature. The collection consists of 40 objective questions (Tables 5, 6, 10−15, 17, 18) and the eight subjective questions of the SNS (Table 16).

We attempt to classify the objective questions in Table 21. The four categories overlap. "Numeration, counting, and hierarchy," which we adopt from Huizinga et al. (2008), contains such questions as reading scales and figuring out appointment times and medication schedules. Probability (risk) accounts for half of the questions. These questions are heavy on ratios, and many involve some arithmetic. Some questions in the numeration category also involve arithmetic.

Table 21. Count of assessment items by numeracy skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
Numeration, counting, hierarchy		4				1	5	3	13
Arithmetic				1	1	1		1	4
Multistep arithmetic				1		1		1	3
Probability	4		8	1	6	1			20
Test: (1) Schwartz et al. 1997; (2) S-TOFHLA, Baker et al. 1999; (3) Lipkus et al. 2001; (4) Estrada et al. 2004; (5) MDIT, Schwartz et al. 2005; (6) ANQ, Apter et al. 2006; (7) TOFHLiD, Gong et al. 2007; (8) DNT, Huizinga et al. 2008 and Cavanaugh et al. 2008.

It is evident from Table 21, as well as the questions themselves, that these numeracy assessments involve elementary, foundational mathematics. There is nothing here that can be construed as algebra, geometry, or calculus.

Concluding Comment (BACK FROM THE MINE)

By definition, journals published in the *Web of Science* (formerly, the *Science Citation Index*) are impact journals. The *WoS* is the library resource that scientists consult to find important papers in their field. A search for the term "numeracy" produces a bibliography of a few hundred papers published in the past ten years, and the bibliography is growing exponentially with a doubling period of ~5 years (Vacher and Chavez 2008). These facts show that numeracy—or more exactly, innumeracy—is getting attention.

It is good news too that so much of the bibliography is composed of an expanding network (Fig. 1) of citation-linked papers in medical journals. With the culture of clinical-based studies and rigorous peer review that characterizes medical research, the implication is that numeracy will be the subject of evidence-based research. Indeed, to us, the most conspicuous feature of the papers in the health numeracy literature was that the statements in them are backed by evidence. The papers not only asserted that numeracy is (astonishingly) low in the population that the medical community deals with; the papers presented evidence. Not only did the papers present evidence; the evidence came from assessment instruments for which the authors included, sometimes in numbing detail, evidence of reliability and validity of the assessment instruments. And, the papers did not simply lament that the rampant innumeracy threatens to affect health care and patient self-management; they presented research evidence of connections. The broader numeracy community can assert with some justification that, in addition to everything else (e.g., Paulos 1988, Steen 2001), widespread innumeracy is bad for our citizens' health.

We believe we see a thematic evolutionary radiation (in the sense of a crude phylogenetic tree) in the collection of health numeracy research papers. The first

papers (notably Williams et al. 1995) sounded the alarm that many patients' functional literacy, including numeracy, is so low that they cannot function in today's health care environment. Schwartz et al. (1997) soon followed with a more specific message: that patients' poor understanding of probability associates with their inability to accurately gauge the benefit of cancer screening. There then developed a branch in which research addressed the correlation of assessed numeracy with other medical outcomes (e.g., Estrada et al. 2004; Apter et al. 2006; Yin et al. 2007; Cavanaugh et al. 2008; Ginde et al. 2008).

Another large branch that developed early has focused on the perception of risk, particularly on finding ways to communicate with innumerate patients about it (Lipkus et al. 2001; Sheridan et al. 2003; Hanoch 2004; Schwartz et al. 2005; Kelly et al. 2007; Schwartz et al. 2007). In addition, a persistent, early branch (e.g., Baker et al. 1999; Fagerlin et al. 2007) seeks to develop assessments of general health numeracy not only to support health numeracy research but also for routine use in clinical settings so that practitioners can identify innumerate patients in order to better communicate with them. And there is a recently twigged branch that investigates numeracy as a phenomenon (Donnelle et al. 2007; Reyna and Brainard 2007, 2008).

It is problematic of course to try to glean a phylogeny from a literature because the actual branches are not mutually exclusive. We interpret branches, though, in order to draw a contrast. On the one hand we have the two branches that represent workarounds: the medical researchers have recognized the pervasiveness of innumeracy and now seek ways of coping with it. Thus the risk-perception branch is finding more effective ways of presenting numerical information to an innumerate public,[24] and the assessment developers are finding ways of mathematically stressing their innumerate subjects less.[25] These themes were noted in the early benchmark papers.[26,27]

---

[24] From Schwartz et al. 2007, p. 661: "We believe our study supports the broader use of simple tabular displays to communicate the benefits and side effects of drugs or other medical interventions to the public. We hope that our findings will encourage medical journal editors, journalists (and their editors), and authors of patient decision aids to routinely use tables to communicate data to the public. Although editors and writers may worry that readers would not understand the data, our findings suggest that most would."

[25] From Fagerlin et al. 2007, p. 679: "[Our study] suggests that replacing the Lipkus and others objective numeracy questionnaire with the SNS might decrease attrition rates for longitudinal studies and increase the likelihood that people will complete questionnaires."

[26] From Williams et al. 1995, p. 9 of 12-page pdf: "… low literacy remains an occult, silent disability. Clinicians must learn to identify these individuals compassionately and overcome communication barriers to ensure that patients with inadequate literacy skills receive high-quality health care."

[27] From Schwartz et al. 1997, abstract: "More effective formats are needed to communicate quantitative information about risks and benefits."

On the other hand, there was a potential branch that appears to have been abandoned. The full last sentence of the paper by Schwartz et al. (1997) is:

> If effective communication of quantitative information is important, *we can either better educate society to improve numeracy* or try to develop communication strategies that overcome innumeracy (emphasis added).

The health numeracy research community is not waiting for society to be better educated; it is researching ways of communicating that compensate for low numeracy. We believe, however, that publications of the health numeracy research community provide resources for educators in the National Numeracy Network to address the problem of low numeracy. One of those resources, we hope, is the set of 48 assessment items we present in this paper.

Acknowledgments

References

American Medical Association Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs. 1999. Health literacy: Report of the Council on Scientific Affairs. *Journal of the American Medical Association* 281: 552–557. http://dx.doi.org/10.1001/jama.281.6.552

Ancker, J. S., and D. Kaufman. 2007. Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association* 14 (6): 713–721. http://dx.doi.org/10.1197/jamia.M2464

Apter, A. J., J. Cheng, D. Small, I. M. Bennett, C. Albert, D. G. Fein, M. George, and S. Van Horne. 2006. Asthma numeracy skill and health literacy. *Journal of Asthma* 43 (9): 705–710. http://dx.doi.org/10.1080/02770900600925585

Arkes, H. R., V. A. Shaffer, and M. A. Medow. 2007. Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making* 27 (2): 189–202. http://dx.doi.org/10.1177/0272989X06297391

Baker, D. W. 2006. The meaning and the measure of health literacy. *Journal of General Internal Medicine* 21 (8): 878–883. http://dx.doi.org/10.1111/j.1525-1497.2006.00540.x

Baker, D. W., M. V. Williams, R. M. Parker, J. A. Gazmararian, and J. Nurss. 1999. Development of a brief test to measure functional health literacy. *Patient Education and Counseling* 38 (1): 33–42. http://dx.doi.org/10.1016/S0738-3991(98)00116-5

Bateman, A.M. 1950. *Economic Mineral Deposits*. New York: Wiley.

Cavanaugh K., M.M. Huizinga, K.A, Wallston, T. Gebretsadik, A. Shintani, D. Davis, R.P. Gregory, L. Fuchs, R. Malone, A. Cherrington, M. Pignone, D. DeWalt, T. Elsay, and R. L. Rothman. 2008. Association of numeracy and diabetes control. *Annals of Internal Medicine* 148 (10): 737–746

Damschroder, L.J., B.J. Zikmund-Fisher, and P.A. Ubel. 2008. Considering adaptation in preference elicitations. *Health Psychology* 17 (3): 394–399. http://dx.doi.org/10.1037/0278-6133.27.3.394

Davis, T. C., M. A. Crouch, S. W. Long,, R. H. Jackson, P. Bates, R. B. George, L. E. Bairnsfather. 1991. Rapid assessment of literacy levels of adult primary care patients. *Family Medicine* 23 (6): 433–435.

Davis, T. C., S. W. Long, R. H. Jackson, E. J. Mayeaux, R. B. George, P. W. Murphy, and M. A. Crouch. 1993. Rapid assessment of adult literacy in medicine: A shortened screening instrument. *Family Medicine* 25 (6): 391–395.

Davis, T. C. and M .S. Wolf. 2004. Health literacy: Implication for family Medicine. *Family Medicine* 36 (8): 595–598.

Donelle, L., L. Hoffman-Goetz, and J. F. Arocha. 2007. Assessing health numeracy among community-dwelling older adults. *Journal of Health Communication* 12 (7): 651–665. http://dx.doi.org/10.1080/10810730701619919

Estrada, C. A., M. Martin-Hryniewicz, B. T. Peek, C. Collins, and J. C. Byrd. 2004. Literacy and numeracy skills and anticoagulation control. *American Journal of the Medical Sciences* 328 (2): 88–93. http://dx.doi.org/10.1097/00000441-200408000-00004

Fagerlin, A., B. J. Zikmund-Fisher, P. A. Ubel, A. Jankovic, H. A. Derry, and D. M. Smith. 2007. Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making* 27 (5): 672–680. http://dx.doi.org/10.1177/0272989X07304449

Fitzgerald, J., M. Funnell, G. Hess, P. Barr, R. Anderson, R. Hiss, and W. Davis. 1998. The reliability and validity of a brief diabetes knowledge test. *Diabetes Care* 21 (5): 706−710. http://dx.doi.org/10.2337/diacare.21.5.706

Ginde, A. A., S. Clark, J. N. Goldstein, and C. A. Camrago, Jr. 2008. Demographic disparities in numeracy among emergency department patients: Evidence from two multicenter studies. *Patient Education and Counseling* 72: 350–356. http://dx.doi.org/10.1016/j.pec.2008.03.012

Golbeck, A. L., C. R. Ahlers-Schmidt, A. M. Paschal, and S. E. Dismuke. 2005. A definition and operational framework for health numeracy. *American Journal of Preventive Medicine* 29: 375–376. http://dx.doi.org/10.1016/j.amepre.2005.06.012

Gong, D. A., J. Y. Lee, R. G. Rozier, B. T. Pahel, J. A. Richman, and W. F. Vann. 2007. Development and testing of the test of functional health literacy in dentistry (TOFHLiD). *Journal of Public Health Dentistry* 67 (2): 105–112. http://dx.doi.org/10.1111/j.1752-7325.2007.00023.x

Gunning, Robert. 1968. *The Technique of Clear Writing*. McGraw-Hill.

Hamm, R. M., D. E. Bard, E. Hsieh, and H. F. Stein. 2007. Contingent or universal approaches to patient deficiencies in health numeracy. *Medical Decision Making* 27 (5): 635–637. http://dx.doi.org/10.1177/0272989X07307516

Hanoch, Y. 2004. Improving doctor-patient understanding of probability in communicating cancer-screening test findings. *Journal of Health Communication* 9: 327–335. http://dx.doi.org/10.1080/10810730490468496

Hibbard, J. H., E. Peters, A. Dixon, and M. Tusler. 2007. Consumer competencies and the use of comparative quality information: It isn't just about literacy. *Medical Care Research and Review* 64 (4): 379–394. http://dx.doi.org/10.1177/1077558 707301630

Hopko, D.R., R. Mahadevan, R.L. Bare and M.K. Hunt. 2003. The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment* 10: 178–182. http://dx.doi.org/10.1177/1073191103010002008

Huizinga, M.M., T.A. Elasy, K.A. Wallston, K. Cavanaugh, D. Davis, R. P. Gregory, L.S. Fuchs, R. Malone, A. Cherrington, D. A. DeWalt, J. Buse, M. Pignone, and R. L. Rothman. 2008. Development and validation of the Diabetes Numeracy Test (DNT). *BMC Health Services Research* 64 (4), Art. No. 96.

Human Resources Development Canada (HRDC). 1997. *Literacy Skills for the Knowledge Society*. Paris: Organisation for Economic Cooperation and Development (OECD) Publications.

Kelly, K. M., K. D. Graves, F. W. K. Harper, J. E. Schmidt, S. L. Dickinson, and M. A. Andrykowski. 2007. Assessing perceptions of cancer risk: Does mode of assessment or numeracy matter? *Cancer Detection and Prevention* 31 (6): 465–473. http://dx.doi.org/10.1016/j.cdp.2007.10.011

Lipkus, I. M., G. Samsa, and B. K. Rimer. 2001. General performance on a numeracy scale among highly educated samples. *Medical Decision Making* 21 (1): 37–44. http://dx.doi.org/10.1177/0272989X0102100105

Lu, M., S.A. Safren, P.R. Skolnik, W.H. Rogers, W. Coady, H. Hardy, and I.B. Wilson. 2008. Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behavior* 12: 86–94. http://dx.doi.org/10.1007/s10461-007-9261-4

Montori, V. M., and R. L. Rothman. 2005. Weakness in numbers - The challenge of numeracy in health care. *Journal of General Internal Medicine* 20 (11): 1071–1072. http://dx.doi.org/10.1111/j.1525-1497.2005.051498.x

Nelson, W., V.F. Reyna, A. Fagerlin, I. Lipkus, and E. Peters. 2008. Clinical implications of numeracy: Theory and practice. *Annals of Behavioral Medicine* 35: 261–274. http://dx.doi.org/10.1007/s12160-008-9037-8

Nielsen-Bohlman L., A.M. Panzer and D.A. Kindig (eds). 2004. Committee on Health Literacy, Board on Neuroscience and Behavioural Health [Institute of Medicine]. *Health Literacy: A Prescription to End Confusion*. Washington DC: National Academies Press.

Parker, R.M., Baker, D.W., Williams, M.V., et al. 1995. The test of functional health literacy in adults – A new instrument for measuring patients literacy skills. *Journal of General Internal Medicine* 10 (10): 537–541. http://dx.doi.org/10.1007/ BF02640361

Paulos, J. A. 1988. *Innumeracy: Mathematical illiteracy and its consequences*. New York NY: Random House.

Peters, E., D. Vastfjall, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert. 2006. Numeracy and decision making. *Psychological Science* 17 (5): 407–413. http://dx.doi.org/10.1111/j.1467-9280.2006.01720.x

Peters, E., J. Hibbard, P. Slovic, and N. Dieckmann. 2007. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs* 26 (3): 741–748. http://dx.doi.org/10.1377/hlthaff.26.3.741

Reyna, V. F., and C. J. Brainerd. 2007. The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences* 17 (2): 147–159. http://dx.doi.org/10.1016/j.lindif.2007.03.010

———. 2008. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences* 18 (1): 89–107. http://dx.doi.org/10.1016/j.lindif.2007.03.011

Richman J., J. Lee, R. Rozier, D. Gong, B. Pahel, and W. Vann Jr. 2007. Evaluation of a word recognition instrument to test health literacy in dentistry: the REALD-99. *Journal of Public Health Dentistry* 67: 99−104. http://dx.doi.org/10.1111/j.1752-7325.2007.00022.x

Rothman, R. L., R. Housam, H. Weiss, D. Davis, R. Gregory, T. Gebretsadik, A. Shintani., and T. Elasy. 2006. Patient understanding of food labels: The role of literacy and numeracy. American Journal of Preventive Medicine 31 (5): 391−398. http://dx.doi.org/10.1016/j.amepre.2006.07.025

Schapira, M. M, K. E. Fletcher, M. A. Gilligan, T. K. King, W. L. Purushottam, and J. M. Neuner. 2008. A framework for health numeracy: How patients use quantitative skills in health care. *Journal of Health Communication* 13: 501–517. http://dx.doi.org/10.1080/10810730802202169

Schield, Milo. 2008. Quantitative literacy and school mathematics: Percentages and fractions. In *Calculation vs. context: Quantitative literacy and its implications for teacher education*, ed. B. L. Madison and L. A. Steen, 87−107. Washington, DC: Mathematical Association of America. http://www.maa.org/Ql/cvc/cvc-087-107.pdf (accessed 12/29/08).

Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch. 1997. The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine* 127 (11): 966–972.

Schwartz, L. M., S. Woloshin, and H. G. Welch. 2005. Can patients interpret health information? An assessment of the medical data interpretation test. *Medical Decision Making* 25 (3): 290–300. http://dx.doi.org/10.1177/0272989X05276860

——— 2007. The Drug Facts Box: Providing consumers with simple tabular data on drug benefit and harm. *Medical Decision Making* 27 (5): 655–662. http://dx.doi.org/10.1177/0272989X07306786

Schwartz, S. R., J. McDowell, and B. Yueh. 2004. Numeracy and the shortcomings of utility assessment in head and neck cancer patients. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck* 26 (5): 401–407.

Sheridan, S. L., M. P. Pignone, and C. L. Lewis. 2003. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *Journal of General Internal Medicine* 18 (11): 884–892. http://dx.doi.org/10.1046/j.1525-1497.2003.21102.x

Smerecnik, C. M. R., and I. Mesters. 2007. Validating the medical data interpretation test in a Dutch population. *Patient Education and Counseling* 68 (3): 287–90. http://dx.doi.org/10.1016/j.pec.2007.06.013

Speros, C. 2004. Health literacy: concept analysis. *Journal of Advanced Nursing* 50 (6): 633–640.

Steen, Lynn Arthur, ed. 2001. *Mathematics and democracy: The case for quantitative literacy*. Princeton NJ: National Council on Education and the Disciplines. http://www.maa.org/ql/mathanddemocracy.html (accessed 12/29/08).

———. 2008. Reflections on Wingspread Workshop. In *Calculation vs. context: Quantitative literacy and its implications for teacher E\education*, ed. B. L. Madison and L. A. Steen, 11–23. Washington, DC: Mathematical Association of America. http://www.maa.org/Ql/cvc/cvc-011-023.pdf (accessed 12/29/08).

Taylor, W. 1953. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly* 30: 415–433.

Tucker, Alan. 2008. Fractions and units in everyday life. In *Calculation vs. context: Quantitative literacy and its implications for teacher E\education*, ed. B. L. Madison and L. A. Steen, 75–86. Washington, DC: Mathematical Association of America. http://www.maa.org/Ql/cvc/cvc-075-086.pdf (accessed 12/29/08).

U.S. Department of Health and Human Services. 2000. *Healthy People 2010*, 2nd ed. With Understanding and Improving Health and Objectives for Improving Health, 2 vols. Washington DC: US Government Printing Office.

Vacher, H. L., and T. Chavez. 2008. Quantitative literacy on the Web of Science, 1: The bibliography and its role in the history of this journal. *Numeracy* 1 (2): Article 2. http://dx.doi.org/10.5038/1936-4660.1.2.2 (accessed 12/29/08).

van Tol-Geerdink, J. J., P. F. M. Stalmeier, E. N. J. T. van Lin, E. C. Schimmel, H. Huizenga, W. A. J. van Daal, and J. W. Leer. 2006. Do prostate cancer patients want to choose their own radiation treatment? *International Journal of Radiation Oncology Biology Physics* 66 (4): 1105–1111. http://dx.doi.org/10.1016/j.ijrobp.2006.07.001

Weiss B.D., M.Z. Mays, W. Martz, K. M. Castro, D. A. DeWalt, M. P. Pignone, J. Mockbee, and F. A. Hale. 2005. Quick assessment of literacy primary care: The newest vital sign. *Annals of Family Medicine* 3: 514–522. http://dx.doi.org/10.1370/afm.405

Wilkinson G. S. 1993. WRAT-33. *Wide Range Achievement Test administration manual*. Wilmington DE: Wide Range Inc.

Williams, M. V., R. M. Parker, D. W. Baker, N. S. Parikh, K. Pitkin, W. C. Coates, and J. R. Nurss. 1995. Inadequate functional health literacy among patients at 2 public hospitals. *Jama-Journal of the American Medical Association* 274 (21): 1677–1682.

Woloshin, S., L. M. Schwartz, M. Moncur, S. Gabriel, and A. N. A. Tosteson. 2001. Assessing values for health: Numeracy matters. *Medical Decision Making* 21 (5): 382–390. http://dx.doi.org/10.1177/0272989X0102100505

Woloshin, S., L. M. Schwartz, and H. G. Welch. 2005. Patients and medical statistics: Interest, confidence and ability. *Journal of General Internal Medicine* 20: 996-1000. http://dx.doi.org/10.1007/s11606-005-0245-7

World Health Organization (WHO). 1998. Division of Health Promotion, Education and Communications Health Education and Health Promotion Unit. *Health Promotion Glossary*. Geneva: World Health Organization.

Yin, H. S., B. P. Dreyer, G. Foltin, L. van Schaick, and A. L. Mendelsohn. 2007. Association of low caregiver health literacy with reported use of nonstandardized dosing instruments and lack of knowledge of weight-based dosing. *Ambulatory Pediatrics* 7 (4): 292–298. http://dx.doi.org/10.1016/j.ambp.2007.04.004

Zarcadoolas, C.A., A. Pleasant and D.S. Greer 2005. Understanding health literacy: An expanded model. *Health Promotion International* 20 (2): 195–203. http://dx.doi.org/10.1093/heapro/dah609

Zikmund-Fisher, B. J., D. M. Smith, P. A. Ubel, and A. Fagerlin. 2007. Validation of the subjective numeracy scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making* 27 (5): 663–671. http://dx.doi.org/10.1177/0272989X07303824

Zikmund-Fisher, B.J., A. Fagerlin, T. Ro. Roberts, H.A. Derry, and P.A. Ubel. 2008. Alternate methods of framing information about medication side effects: Incremental risk versus total risk of occurrence. *Journal of Health Communication* 13 (2): 107–124. http://dx.doi.org/10.1080/10810730701854011

Appendix
Applications of Health Numeracy Assessments
(arranged chronologically)

Sheridan et al. (2003)
- The study: How to communicate potential benefits and harms to patients: number needed to treat (NNT), absolute risk reduction (ARR), relative risk reduction (RRR) or a combination (COMBO). Participants were given written information about the baseline risk of a hypothetical disease and were asked to decide which of two drug treatments would provide more benefit and, further, to calculate the effect of one of those drug treatments on the given baseline risk of disease. Risk information was given in one of the four investigated risk formats, selected randomly.
- Numeracy assessment instrument: Three-item Schwartz et al. (1997) test.
- Participants: 358 patients, ages 50-80, recruited from a university internal medicine clinic.
- Numeracy findings (see Table 9B): Overall, RRR was the more successful format (60% correct, compared to 43% for COMBO, 42% for ARR, and 30% for NNT). With respect to numeracy: 88% of the 2% of the participants who got all 3 questions correct stated correctly which treatment provided more benefit (compared to 63% of the 28% who got 2 numeracy questions correct, and 35% of the 71% who got 1 or 0 numeracy questions correct). Results for the calculation were: 50%, 30% and 5% correct for the participants with 3, 2, and ≤1 numeracy items correct, respectively.

S. Schwartz et al. (2004)
- The study: Connecting numeracy to utility assessment (SG and TTO) and a one-item questionnaire to assess overall ("global") quality of life (QOL).
- Numeracy assessment instrument: Three-item Schwartz et al. (1997) test via Woloshin et al. (2001).
- Participants: 18 patients who had recently been diagnosed with head and neck cancer.
- Numeracy findings: Half of the patients were numerate (answered 2-3 questions correctly). The numerate patients gave more consistent responses on the other assessments, and those assessments agreed more with clinician-rated functioning. Paper concludes: "QOL evaluation through utility assessment may provide inaccurate and contradictory data about patient functioning for nonnumerate patients. This may confound QOL assessment when interpreting utility data" (p. 401).

Van Tol-Geerdink et al. (2006)
- The study: Which patients want to be involved in choosing the radiation dose in their conformal radiation therapy, and which patients want to leave the choice to

the doctor? The patients were informed of the benefits and risks associated with two choices by use of side-by-side pie charts ("decision aids") comparing the choices for each of several outcomes and side effects. Numeracy was one of several independent variables.

- Numeracy assessment instrument: Schwartz et al. (1997) test.
- Participants: 140 prostate cancer patients at two hospitals in the Netherlands.
- Numeracy findings (see Table 9B): Patients with high numeracy were more likely to accept the option to choose: 72 of the 83 numerate patients (87%) opted to choose; 40 of the 57 nonnumerate patients (70%) opted to choose.

Peters et al. (2006)

- The study: Four studies examining the roles of numeracy (ability to comprehend and transform probability numbers) and affect in decision making.
- Assessment instrument: Lipkus et al. (2001)
- Participants: 100 participants were recruited using a campus newspaper for one of the studies. There were 46 students from a psychology course in two of the other studies, and 171 volunteers from the psychology department in the fourth.
- Numeracy findings (see Table 11): "… highly numerate individuals (are) more likely to retrieve and use appropriate numerical principles, thus making themselves less susceptible to framing effects, compared with less numerate individuals. In addition, the highly numerate tended to draw different (generally stronger or more precise) affective meaning from numbers and numerical comparison, and their affective responses were more precise." (p. 407).

Yin et al. (2007)

- The study: Is low health literacy of caregivers (e.g., parents) associated with incorrect medication dosing of young children? Independent variables: health literacy including numeracy. Dependent variables: (1) use of nonstandardized dosing instrument – subjects were asked "*What do you use most of the time at home to give your child his/her liquid medicine?*" and shown photographs of kitchen teaspoon, kitchen tablespoon, dosing spoon, measuring spoon, dosing cup, dropper and syringe; .(2) lack of knowledge of weight-based dosing – "*From the following list, what is the most important characteristic of a child when deciding what dose of medicine to give*?" – gender, age, height, weight, and how the child is feeling.
- Assessment instrument: TOFHLA (Williams et al. 1995).
- Participants: 292 caregivers recruited from an urban pediatric emergency room.
- Numeracy findings: 23% of the participants reported using nonstandardized dosing instruments, and 68% were unaware that dosages are based on weight. "Overall health literacy, reading comprehension and numeracy were all associated with both dependent variables" (p. 292).

Kelly et al. (2007)

- The study: An assessment of cancer risk perception. Participants were asked four risk perception items in the context of colon cancer: (1) estimate the likelihood of developing colon cancer in your lifetime (personal percentage risk

perception), (2) estimate the percentage likelihood of a typical woman your age in developing colon cancer in her lifetime (population risk); (3) compare your lifetime risk of developing colon cancer to other women your age using a five-point scale from "much below average" to "much above average" (comparative risk); and (4), yes or no, will you develop colon cancer in your lifetime (personal binary risk)?

- Numeracy assessment instrument: Schwartz et al. (1997) test.
- Participants: 457 healthy women (mean age 61.3) either participating in a cancer-screening program or recruited from through newspaper advertisements.
- Numeracy findings: Although the study found that mode of assessment (telephone, mail, web-based) does not matter to assessing perceived colon cancer risk, it did find that numeracy matters. "Specifically women with lower numeracy tended to yield higher percentage estimates of cancer risk. On the other hand, as hypothesized, numeracy did not appear to influence verbal estimates of perceived risk. Similar risk estimates were reported for the binary and comparative risk items regardless of a woman's level of numeracy. These results suggest individual differences in numeracy may be a source of error in the assessment of cancer risk perceptions when numerically based risk perception items are used" (p. 471).

Hibbard et al. (2007)
- The study: Contribution of health literacy, numeracy and patient activation (knowledge, skill and confidence for managing one's own health and health care) to patients' comprehension of comparative health care performance reports and in making an informed choice.
- Numeracy assessment instrument: Lipkus et al. (2006) plus "four more difficult items…. The items assess people's ability to understand risk magnitudes, to calculate percentages, to convert proportions to percentages, and probabilities to proportions. An example question was: Which is a higher risk: 1 in 10, 1 in 100, or 1 in 1000?" In addition, the study assessed participants' comprehension as a dependent variable. The 13-item comprehension assessment, which is included in the appendix of the paper, asks the participants to read comparative data displays and answer 3-4 questions requiring a conclusion: for example, which hospital has the highest death rate for patients being treated for heart failure?
- Participants: 303 employed-age adults recruited using posted fliers and newspaper advertisements
- Numeracy findings: Scores on the 15-item numeracy test ranged from 0 to 15, and had a mean of 9.3. Numeracy was found to be the strongest predictor of comprehension, and could be compensated for, to some extent, by high activation. The authors concluded: "… the main focus on consumer competencies for making health care choices has been on health literacy. While heath literacy is an important determinant of use of information, it does not tell the whole story, and in fact is not even the most predictive of the three competency areas explored in this study. Efforts to support informed patient choices will be more successful if they focus on addressing patients' numeracy

and literacy skills, and activation levels, than if they only focus on addressing health literacy" (p.389).

Arkes et al. (2007)

- The study: How do patients react to a physician who bases a decision on a computer-assisted diagnostic support system (DSS)? The participants were undergraduate students. The study involved four separate experiments in which students read short scenarios describing a physician-patient interaction and then rated target independent variables such as diagnostic ability of the physician and thoroughness of the exam. In one of the experiments, for example, the scenario had the student going to the physician after turning an ankle in a softball game. In one version, the physician orders an X-ray; in a second version, the physician enters the information into a DSS and uses it to make the decision to order the X-ray; in the third version, the physician orders the X-ray after using a "prestigious" DSS (i.e., the physician tells the patient that the computer program was "developed at the prestigious Mayo Clinic, one of the nation's premier medical facilities"). Similar to other studies, the values of the dependent variables were highest for the no-aid scenario and least for the prestigious-aid scenario.
- Assessment instrument: The three general numeracy questions from Lipkus et al. (2001) (i.e., the three questions from Schwartz et al. 1997, with the 2-sided coin traded out for a 6-sided die).
- Participants: 347 undergraduate students who participated as part of their course.
- Numeracy findings. The average score on the 3-item numeracy assessment was 1.83. Contrary to hypothesis, numeracy did not significantly affect the evaluation of the physician.

Donnelle et al. (2007)

- The study: Used a variety of assessment tools to examine relationships between health numeracy skills and prose health literacy, education, and math anxiety.
- Assessment instruments: S-TOFHLA to measure health literacy (prose and numeracy), the three Schwartz et al. (1997) general-context numeracy items, and the eight Lipkus et al. (2001) health-context numeracy items. They measured math anxiety by Abbreviated Math Anxiety Scale (AMAS) of Hopko et al. (2003).
- Participants: 140 independently functioning older adults (>50 years old) recruited from communities in southern Ontario. "In effect, these results are illustrative of a 'well' rather than a health-compromised group of older adults" (p. 662).
- Numeracy findings (see Table 9B): For the eight health-context numeracy questions from Lipkus et al., the mean score was 5.9. "Approximately 36% of the variation in general-context numeracy and 26% of the variation in health-context numeracy were explained by prose health-literacy skill, math anxiety and attained education" (p. 651).
- Moreover: The numeracy results depended on the numeracy assessment

instrument: numeracy measured by STOFHLA was higher than numeracy measured by the health-context items, which was higher than the numeracy measured by the general-context items. The authors remarked on "the lack of a single, comprehensive numeracy measurement tool appropriate for health-based research." The three assessment instruments "may not be representative of the full construct of health numeracy." As a further problem, they pointed out a lack of "consistency in defining health numeracy. Without agreement on what this construct represents, it is difficult to operationalize numeracy accurately for the purposes of research and, furthermore, to compare research findings accurately" (p. 663).

Ginde et al. (2008)
- The study: Two cross-sectional studies to estimate the prevalence of limited numeracy amongst patients at emergency departments and evaluate for demographic disparities.
- Assessment instrument: A four-item general numeracy assessment that combined the three general numeracy items of Lipkus et al. (2001) and the three original questions from Schwartz et al. (1997) (i.e., question 1 from each of those two tests, plus questions 2 and 3, which both tests used).
- Participants: 897 patients took the numeracy test, 536 in the first study and 361 in the second study. Participants for the first study were from 812 consecutive patients who presented for acute asthma at 26 EDs in 17 US states (770 were eligible, and 590 were enrolled in the program). Participants for the second study were from 1080 patients at 4 EDs in one state (640 were eligible, and 473 were enrolled.
- Numeracy findings (see Table 9A): For the second study, 20% had no correct answers, 22% had one correct, 29% had two correct, 17% had three correct, and 11% had all four correct. Limited numeracy was more prevalent among ED patients than in the general population, and there were significant disparities. Older age, minority race/ethnicity, and limited education are independently associated with limited numeracy. "Practitioners should pay particular attention to numeracy in acute care settings, such as the ED, where patients appear to be at higher risk for limited numeracy and the consequences of ineffective communication may be higher" (p. 354).