

Clinical Implications of Numeracy: Theory and Practice

Wendy Nelson, Ph.D.,¹ Valerie F. Reyna, Ph.D.,² Angela Fagerlin, Ph.D.,³ Isaac Lipkus, Ph.D.,⁴

Ellen Peters, Ph.D.⁵

¹National Cancer Institute; ²Cornell University and National Cancer Institute; ³Ann Arbor VA

HSR&D Center for Excellence Division of General Internal Medicine and University of

Michigan Center for Behavioral and Decision Sciences in Medicine; ⁴Duke University

⁵Decision Research, Eugene, Oregon, and University of Oregon

Corresponding author: Wendy Nelson, Ph.D., Basic and Biobehavioral Research Branch, BRP,

DCCPS, NCI, 6130 Executive Blvd., Bethesda, MD 20892. E-mail: nelsonw@mail.nih.gov

Acknowledgements: This work was based on a National Cancer Institute-sponsored symposium – *Numeracy: A Critical (and Often Overlooked) Competence for Health Decision Making* - presented at the Society of Behavioral Medicine Annual Meeting, Washington, D.C., March 22, 2007. Dr. Reyna is supported by grants from the National Cancer Institute (R13CA126359) and the National Institute of Mental Health (MH-061211). Dr. Fagerlin is supported by an MREP early career award from the U.S. Department of Veterans Affairs. Dr. Lipkus is supported by The Foundation for Informed Medical Decision Making. Dr. Peters is supported by a grant from the National Science Foundation (SES-0517770). We thank Nathan Dieckmann for his helpful literature review.

Running head: Clinical implications of numeracy

In press, *Annals of Behavioral Medicine*.

ABSTRACT

Background: Low numeracy is pervasive and constrains informed patient choice, reduces medication compliance, limits access to treatments, impairs risk communication, and affects medical outcomes; therefore, it is incumbent upon providers to minimize its adverse effects.

Purpose: We provide an overview of research on health numeracy and discuss its implications in clinical contexts.

Conclusions: Low numeracy cannot be reliably inferred on the basis of patients' education, intelligence, or other observable characteristics. Objective and subjective assessments of numeracy are available in short forms, and could be used to tailor health communication. Low scorers on these assessments are subject to cognitive biases, irrelevant cues (e.g., mood), and sharper temporal discounting. Because prevention of the leading causes of death (e.g., cancer and cardiovascular disease) depends on taking action now to prevent serious consequences later, those low in numeracy are likely to require more explanation of risk to engage in prevention behaviors. Visual displays can be used to make numerical relations more transparent, and different types of displays have different effects (e.g., greater risk avoidance). Ironically, superior quantitative processing seems to be achieved by focusing on qualitative gist and affective meaning, which has important implications for empowering patients to take advantage of the evidence in evidence-based medicine.

INTRODUCTION

In this era of evidence-based medicine, patients and their health care providers are inundated with numbers. Whether it is direct-to-consumer advertising proclaiming that a clinical study of the latest cholesterol-lowering drug found 42% fewer deaths from heart attack, or a medication prescription instructing patients to take half a 10 mg. tablet twice a day, or a newspaper article summarizing changes in cancer incidence, people cannot seem to escape having to deal with numbers in order to make informed decisions about health. Like literacy, numeracy is a necessity for daily life. It is especially critical in the health domain, where understanding or not understanding what numbers mean may have life-altering consequences. For example, numerical competence is needed to understand and weigh the risks and benefits of treatment, to decipher survival and mortality curves, and to navigate medical insurance forms and informed consent documents. Now that more and more people are seeking medical information from the Internet, numeracy is especially critical. The National Cancer Institute's 2005 Health Information National Trends Survey, which comprises a nationally representative sample of the general adult population, found that nearly 50% of all people seeking information about cancer went to the Internet as their *first* source of information, whereas only 23% of people initially sought information from their health care provider (1). On the one hand, this is encouraging news: people are becoming more informed health care consumers. On the other hand, how do we know that people understand the numerical information they see online? There are now Web sites that allow you to calculate your risk of developing various types of cancer, diabetes, cardiovascular disease, and osteoporosis. For example, if a woman wants to know her risk of developing breast cancer, she can search for "breast cancer risk," click on the link for the National Cancer Institute Breast Cancer Risk Assessment Tool, answer six simple questions, and

find out her five-year and lifetime risk of developing breast cancer. But exactly what do these numbers mean? Is a 9.2% lifetime risk of developing breast cancer reassuring or worrisome? Herein rests the challenge of numeracy for behavioral scientists and clinicians: how to make numbers transparent and comprehensible so that individuals can make intelligent, reasoned decisions about their health.

Numeracy defined in the broadest sense is the ability to comprehend, use, and attach meaning to numbers. Also referred to as quantitative literacy, numeracy is often subsumed under the broader construct of literacy. Both the National Adult Literacy Survey (2) and the National Assessment of Adult Literacy (3) operationalize literacy into prose, document, and quantitative domains. Whereas prose literacy involves the ability to extract information from prose texts, both document and quantitative literacy require facility with numbers. Document literacy involves the skills needed to search for and use text and numeric information contained within documents, such as computing calories from nutrition labels or interpreting tables and graphs. Quantitative literacy involves the ability to apply arithmetic operations using numbers in printed materials. Balancing a checkbook, computing tax on an order form, and calculating the cost of an item that is reduced by 25% are all examples of quantitative literacy tasks. More specifically, health literacy refers to the “capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions (4).” Health literacy and numeracy encompass those cognitive skills that enable individuals to act on health information in ways that promote good health (5,6). More recently, researchers and clinicians have focused on distinguishing health numeracy from health literacy. Golbeck and colleagues (7) define health numeracy as “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.” Ancker and Kaufman (8) suggest that we look beyond individual quantitative skills and conceive of health numeracy as the “productive use of quantitative health information.” In their conceptualization, health numeracy goes beyond the individual’s basic quantitative skills (i.e., computation, estimation, statistical literacy) to encompass the ability to use numeric data contained in documents and graphics, and oral communication skills.

While these descriptions of numeracy are helpful, and represent current thinking, it is important to distinguish such descriptions from empirically confirmed dimensions of cognition. In particular, components of these definitions (e.g., productive use of quantitative health information) that are attributed to “numeracy” could be due instead to education or general intelligence. However, research has demonstrated that numeracy has effects on decisions that are independent of education (9), verbal intelligence (10), and health literacy (11), thereby supporting the validity of numeracy as a separate construct (9). In this paper we provide an overview of health numeracy and discuss its implications for providers and patients in clinical contexts.

WHAT WE KNOW ABOUT NUMERACY

The 2003 National Assessment of Adult Literacy, which surveyed a nationally representative sample of adults, found that 22% of adults performed at a below basic quantitative literacy skill level, 66% performed at a basic or intermediate skill level, and 13% performed at a proficient level (12). Individuals at the lowest skill level could not perform simple, one-step arithmetic operations. Equally dismal functional health literacy findings were reported by Williams and colleagues: 22.0% to 61.7% of patients attending two urban, public hospitals had

inadequate or marginal functional health literacy (13). A substantial number of patients did not understand instructions on medication bottles, could not determine when their next appointment was scheduled, and could not determine their eligibility for financial assistance. In short, these patients were unable to perform the most basic tasks required to function in the health care environment. Not surprisingly, average health literacy was correlated with level of educational attainment (3). However, one should not presume a high degree of numeracy based on years of education or an advanced educational degree. For example, Rothman and colleagues (14) studied comprehension of food labels among 200 generally well educated primary care patients. Although 77% of patients had at least high school-level literacy skills (as assessed by the REALM) (15), 63% of patients lacked 9th grade mathematics skill (as assessed by the WRAT-3) (16). In another study involving 357 university clinic patients, more than half of the participants who had at least some college education answered only one or no questions correctly on the three-item general numeracy assessment developed by Schwartz and colleagues (17,18). Studies have shown that physicians (19), medical students (20), and well educated laypersons (21) often have difficulty performing relatively simple arithmetic calculations and comprehending numeric risk estimates, regardless of whether they are expressed numerically (e.g., as percentages) or graphically (e.g., as survival or mortality curves). For example, Yamigishi (22) found that undergraduates had difficulty interpreting simple mortality data: students rated cancer that killed 1,286 out of 10,000 people as riskier than cancer that killed 24.14 out of 100 people. Lipkus et al. (21) found that well educated individuals had difficulty answering simple questions regarding probability, percentages, and proportions: only 20% of subjects were able to convert “1 in 1000” to a percentage. Finally, Reyna et al. (23) showed how physicians and patients had great difficulty understanding genetic risk and were prone to similar errors, despite vast differences in

medical knowledge. For example, both groups confused the probability of disease given a genetic mutation with the converse probability.

Particularly problematic for physicians and patients is understanding risk estimates used to summarize the risks and benefits of medical interventions: relative risk reduction, absolute risk reduction, and numbers needed to treat. This is worrisome because studies have shown that treatment preferences may be influenced by which of these summary statistics is used (19, 24-30). For example, one study found that physicians were more inclined to treat patients when the results of a clinical trial were reported in terms of relative risk reduction than when the same results were reported in terms of absolute risk reduction (19). In this study, physicians were presented with actual clinical trial outcome data that were summarized as either absolute or relative changes in mortality. Ninety-seven of the 235 (41%) surveyed physicians indicated that they would be more inclined to treat with the study drug when results were reported as a relative change than when results were reported in terms of absolute change. Laypersons' decisions about medical treatment appear to be similarly influenced by the format in which benefits are presented. One study found that the lay public was significantly more receptive to cancer screening tests when the benefits of screening were expressed as a relative risk reduction than when benefits were presented in terms of absolute risk reduction or numbers needed to screen (30).

Low numeracy has been associated with a number of undesirable health outcomes (31,32), including self-reported poor health (33), health disparities (34), poor health knowledge and disease self-management skills (35), and choosing lower-quality health options (11). In other studies, individuals with inadequate health literacy, as assessed by the Test of Functional Health Literacy in Adults (TOFHLA) (36) or the S-TOFHLA (37), were more likely to be

hospitalized and less likely to use clinical preventive services than those with adequate health literacy (38,39). Low numeracy has also been associated with inferior disease management. For example, in a prospective cohort study of patients taking warfarin, low numeracy was significantly associated with poor anticoagulation control (40). In a study of asthma patients prescribed inhaled steroids, low numeracy was associated with a history of hospitalizations and emergency room visits for asthma (41). Although the ability to understand and interpret nutrition labels is not a health outcome *per se*, it has important implications for health and disease management. Rothman and colleagues (14) found that comprehension of nutrition labels was significantly associated with numeracy skill - as assessed by the WRAT (16) - and that this association held after adjusting for education and income. Similarly, medication refill adherence is not a health outcome *per se*, but is a critical component of disease management. In a prospective study of Medicare patients with cardiovascular-related diseases, low health literacy, as assessed by the S-TOFHLA, was associated with poor medication refill adherence, although this relationship did not achieve statistical significance (42).

Numeracy also has important health policy implications in that utilities are used to incorporate patient values into pharmacoeconomic analyses and health policy. A utility score quantifies a patient's value or preference for a particular health state (43). Utility assessment techniques, which include the standard gamble and time trade-off, are inherently quantitative tasks that require comparing probabilities. While assessment techniques have been widely used to assess health-related quality of life for a variety of chronic and acute health states (44), their validity for low-numerate individuals is questionable. Woloshin and colleagues (45) asked 96 women how they valued their current state of health by using the standard gamble and time trade-off techniques. One would expect that a higher utility for current health would correlate

with better health. However, they discovered that among low-numerate women, the correlation between utility for current health and self-reported health was in the wrong direction, whereas for high-numerate women, the correlation between utility for current health and self-reported health was in the expected direction. Schwartz and colleagues (46) asked a sample of head and neck cancer patients to complete three utility assessment exercises: the standard gamble, time trade-off, and rating scale. They found that correlations between the three scores were stronger for the numerate than the non-numerate. These studies suggest that utilities obtained from low-numerate individuals may have limited validity.

Recognizing the potential medical and psychosocial harms that may result from low health literacy and numeracy specifically, it is incumbent upon clinicians and researchers to take steps to minimize adverse effects of low numeracy on health outcomes. Perhaps the first thing clinicians need to do is realize that it is not obvious which individuals are low literate and low numerate. Appearances may be deceiving, and national statistics indicate that low numeracy is pervasive (9). For example, clinicians often assume that people are functionally literate because of their physical appearance, level of education, or socioeconomic status, and consequently overestimate literacy skills (47-49). A number of studies have demonstrated that educational attainment is not necessarily a reliable indicator of health literacy (50-53); in fact, the final grade of formal schooling completed is often considerably higher than the individual's functional literacy level (54). We would expect similar assumptions with respect to numeracy.

Furthermore, asking patients to assess their literacy skill is often nonproductive, as data from the National Adult Literacy Survey revealed that most of the adults who performed at the lowest literacy level reported that they could read "well" or "very well" (2). Similarly, self-reported numeracy is not necessarily a reliable predictor of objective numerical performance (55). Given

that low numeracy is common, may not be apparent from observable characteristics, and cannot always be reliably assessed simply by asking, we offer the following recommendations for ways to recognize and remediate low numeracy in the health care setting, based on empirical evidence.

CLINICAL IMPLICATIONS: HOW TO PROMOTE UNDERSTANDING OF NUMBERS

Assessing Numeracy

There are currently two ways to assess adult numeracy. The first method involves evaluating people's "objective" numeracy skills – that is, their ability to perform basic arithmetic operations, to solve problems involving frequency, probability, and percentages, and to interpret health information. The second method involves assessing "subjective" numeracy skills – that is, how confident and comfortable people feel about their numerical ability.

A widely used instrument for assessing objective numerical competence is an 11-item Numeracy Scale based on three questions developed by Schwartz et al. (17) and expanded by Lipkus et al. (21). The original three-item instrument is a content-free numeracy assessment that covers basic familiarity with probability and the ability to convert a percentage to a proportion, and a proportion to a percentage. The instrument has moderate internal consistency, with Cronbach's alpha scores ranging from 0.57 to 0.63 in three separate studies (21). In two studies, higher numeracy scores were associated with greater accuracy: Schwartz et al. (17) reported that women with higher scores were better able to assess the benefit of screening mammography than women with lower scores, and Sheridan and colleagues (18) found that patients with higher scores were better able to interpret treatment benefits than patients with lower numeracy scores. The expanded instrument (21), which is specific to health, assesses the ability to compare risk magnitude, convert percentages to proportions, convert proportions to percentages, and convert

probabilities to proportions. For example, the following question assesses understanding of risk magnitude:

Which of the following numbers represents the biggest risk of getting a disease?

__ 1 in 100, __ 1 in 1000, __ 1 in 10

The scale is easy to score: score 0 for incorrect or missing answers and score 1 for correct answers. It has adequate internal consistency, with Cronbach's alpha scores ranging from 0.70 to 0.75 in three separate studies (21). 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.

In contrast to the Lipkus et al. Numeracy Scale (21), which assesses basic arithmetic and statistical skill, several instruments measure functional health numeracy – that is, the ability to understand and act on numerical health information and thereby function effectively in the health care environment. The TOFHLA (36) is a two-part test consisting of a 50-item reading comprehension section composed of three prose passages and a 17-item numeracy section. It can take up to 22 minutes to administer. The numeracy items pertain to practical tasks, such as understanding labeled prescription vials, interpreting the results of a blood glucose test, and understanding clinic appointment slips and financial assistance information. The number of correct numeracy responses is multiplied by 2.941 to yield a score between 0 and 50. This score is added to the reading comprehension score to yield a final score ranging from 0 to 100. The TOFHLA has excellent split-half and internal consistency reliability (0.92 and 0.98, respectively) and demonstrates good correlation with the REALM and WRAT-R (0.84 and 0.74, respectively). A shortened version of the TOFHLA, the S-TOFHLA (37), contains a 36-item reading

comprehension section made up of two prose passages, and four numeracy items that involve reading a label on a prescription bottle and understanding how and when to take medication, interpreting blood test results, and identifying an appointment date on a clinic appointment slip. The S-TOFHLA can be administered in 12 minutes. Each correct numeracy item is awarded 7 points and each correct reading comprehension item is assigned a score of 2 points, for a total score ranging from 0 to 100. The S-TOFHLA has 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 (37). Both the TOFHLA and S-TOFHLA are available in English and Spanish.

The Medical Data Interpretation Test (56) is an 18-item measure of the ability to interpret medical statistics and understand concepts related to disease risk and risk reduction. Like the TOFHLA, the Medical Data Interpretation Test includes questions that reflect the type of health information that people ordinarily encounter in their daily lives. For instance, two questions pertain to assessing the benefit associated with a hypothetical drug designed to prevent strokes. Items for the test were developed based on reviews of the medical literature and input from experts in statistics, education, cognitive psychology, education, and evidence-based medicine. The Medical Data Interpretation Test has good test-retest repeatability (Pearson $r = 0.67$) and good internal consistency (Cronbach's $\alpha = 0.71$).

Another functional health numeracy screening measure is The Newest Vital Sign (57). This test requires a person to answer six questions based on data contained in a nutrition label from a pint of ice cream. For example, the nutrition label states that the total carbohydrate content in a half-cup serving of ice cream is 30g. People are asked how much ice cream they

could eat if they were permitted to consume 60g of carbohydrates. The Newest Vital Sign has good reliability and validity, for both the English version (Cronbach's alpha = 0.76) and Spanish version (Cronbach's alpha = 0.69), and requires only three minutes to administer. Like The Newest Vital Sign, the Nutrition Label Survey (14) is a 26-question instrument that assesses the ability to interpret actual food labels. Half of the survey involves interpreting food labels (e.g., compute the number of grams of dietary fiber in a serving of candy) and half of the survey involves selecting which of two foods has more or less of a particular ingredient (e.g., choose the product which contains fewer calories from fat: 1 slice of Arnold Hearty Classics 100% Whole Wheat Bread or 4 ounces of Regular Mueller's Spaghetti). Although the Nutrition Label Survey was designed as a survey and not a scale, it has good internal reliability (Kuder-Richardson-20 = 0.87) (personal communication, Russell Rothman). Other researchers have developed numeracy assessment instruments for specific diseases. For example, Estrada and colleagues (40,58) added items specific to warfarin therapy to the three-item numeracy measure developed by Schwartz et al. (17) in order to assess the numeracy of patients on anticoagulation therapy. Other numeracy measures have been developed for patients with asthma (41) and diabetes (59).

The second method for gauging numeracy involves assessing one's subjective numerical ability. The Subjective Numeracy Scale (SNS) developed by Fagerlin and colleagues (60,61) was conceived as a less stressful alternative to objective numeracy measures. Because many people find mathematical aptitude tests aversive, the SNS was designed to predict level of numeracy based on self-assessment of quantitative ability. The SNS contains four items that measure people's perception of their numerical ability and four items that pertain to preferences for numeric information presentation. An example of a question that addresses preference for display of numeric information is as follows:

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)?

The scale demonstrates good reliability (Cronbach’s alpha = 0.82) and is significantly correlated with the Lipkus et al. objective numeracy scale ($r = 0.63 - 0.68$) (60,61). In a head-to-head comparison, the SNS took significantly less time to complete and was viewed as significantly less stressful and frustrating than the Lipkus et al. objective measure (60). Test-retest data are not available for the SNS. The major weakness of the SNS is that it measures perception of ability rather than actual ability.

The STAT-Confidence scale developed by Woloshin and colleagues (62) is a three-item subjective measure that assesses one’s confidence in the ability to understand medical statistics. The scale consists of the following statements: (1) In general, how easy or hard do you find it to understand medical statistics?; (2) I am confident that I can make sense of medical statistics; and (3) I feel like I do not know how to interpret medical statistics. The scale has good psychometric properties (test-retest repeatability $r = 0.62$; Cronbach’s alpha = 0.78). However, unlike Fagerlin et al.’s SNS, the STAT-confidence scale showed only a weak correlation ($r = 0.15$; $p = 0.04$) with an objective measure of numeracy, the Medical Data Interpretation Test (52).

Both the objective and subjective numeracy scales have advantages and disadvantages. For example, while subjective numeracy assessments may be less threatening than objective measures that require arithmetic calculations or data interpretation, an objective numeracy assessment may be more likely to yield a truer estimate of a person’s numerical competence. Any of these brief instruments could be administered to patients prior to an office visit in order to focus the clinical conversation at an appropriate level. Additionally, as patient decision aids

become more commonly used in clinical practice, it will be important to assess numeracy to know whether a particular decision aid is appropriate for a given patient to use.

Making Numbers More Transparent With Visual Displays

Even for low-numerate individuals it may be possible to make clinically relevant numbers more accessible. That is, it may be possible to improve how accurately people understand risk, while at the same time reduce cognitive burden by presenting numeric information in the most efficient and comprehensible manner. Research has shown that including risk statistics in a graphic format can improve understanding (63) and significantly affect decision making (64). Graphs and other visual displays, such as cartoons and films, are now frequently recommended and used in clinical settings as adjuncts to enhance comprehension of statistical information for purposes of risk communication and medical decision making (65,66). Graphic displays can be valuable in helping individuals detect data patterns (e.g., linear trends), perform rudimentary arithmetic operations (e.g., comparisons), and see whole-to-part relationships, such as those that involve conditional probabilities, ratios, and proportions – numerical concepts that are difficult for many to grasp (66-68). As such, graphic displays can enhance the meaning – or gist – of statistical data (68-70)

Some general recommendations can be made with respect to the types of graphic displays that may be best for communicating risk information, particularly for those who are least numerate. Several excellent review papers have reported which graphic formats appear to be most effective for communicating numerical information for different clinical purposes (65,66,71). For example, a line graph is typically the best choice to illustrate trends (e.g., the effectiveness of a drug over time) whereas a bar graph is most effective at showing how rates of

adverse events for different medical treatments compare. Pie charts are useful for judging relative proportions (72). In general, people intuitively grasp the use of height in visual displays such as bar graphs and risk ladders to signify a quantity (e.g., level of risk). People seem to readily comprehend that events located higher on a risk ladder or vertical graph convey a greater quantity than events located toward the bottom (71,73,74). Pictographs represented by figure icons showing the number in a population affected by some event can be used to illustrate magnitude and convey the notion of randomness.

Although using graphs to communicate risk information can improve patients' knowledge and understanding, the precise formatting of the graph can have a tremendous impact on people's ability to interpret information. For instance, when line graphs are used to present survival and mortality rates, people often fail to adjust their risk perception to account for the time frame represented. Zikmund-Fisher and colleagues found that people tend to perceive greater risk and larger differences in treatment effectiveness when a longer period of time (15 years) is represented than when the same mortality risk is represented over a shorter period of time (5 years) (75). Some people have difficulty interpreting pie graphs because they are unable to determine the exact proportion being represented. When using pictographs, it is best to display figures in a systematic, rather than random, fashion. Random displays tend to decrease the precision of risk estimates because relative magnitudes are not easily judged (76). And finally, when using visual displays to represent risk, it is important to consider that visual displays that only emphasize the numerator (e.g., showing only affected individuals) tend to increase risk avoidant behaviors, while those that highlight both the numerator and denominator tend to decrease risk avoidant behavior (68,77,78). Therefore, when using graphs, the appropriate format should be selected in order to avoid bias and improve comprehension.

Preferences for Information Presentation

How should clinicians and health care researchers choose a format for presenting risk information? One could select the visual display that has been shown empirically to promote the greatest understanding or one could ask the patient if he or she had a preference for a particular type of format. In a qualitative study of risk communication formats, Schapira and colleagues found that people expressed definite preferences for both the graphic format and time frame used to present risk information (79). Differences in preferences were noted for better-educated individuals compared to less-educated individuals, with the less-educated group preferring simpler visual displays. Whereas some people prefer bar graphs, others may prefer line graphs or pie charts. Therefore, one option might be to present an array of graphic formats and ask people which they prefer (commonly available computer programs readily convert data into different graphic formats). But what if an individual prefers a graphic format that research suggests is not the most optimal way to communicate numerical information? For example, it has been demonstrated that although people are better at interpreting two-dimensional bar graphs, they prefer three-dimensional bar graphs (80,81). Fagerlin and colleagues tested how well people understood five types of graphs - bar graphs, pictographs, modified pictographs, pie graphs, and modified pie graphs – and found that while bar graphs were preferred over the other formats, they consistently yielded lower scores on gist and verbatim comprehension (69,82). Thus, some people may prefer graphic formats that hinder their ability to interpret data accurately. Nevertheless, there may be situations in which it is advantageous to allow individuals to select their preferred format. For example, people might be more inclined to use

patient decision aids if graphic formats were tailored to their preferences for numeric presentation.

Research unequivocally supports the conclusion that information presentation formats should not be based on the intuition of clinicians or researchers, since doing so may not maximize comprehension of numeric information. A commonly held misconception is that there is no such thing as too much information (68). People reason that if some information is good, then more information must be better. But simply providing all available information does not necessarily enhance comprehension and decision making. In fact, providing more complete information without close attention to content and format can result in poorer comprehension and inferior decision making (69,83). This may be true even when decision makers express a preference for additional information, since people frequently fail to distinguish relevant from irrelevant information, and often end up pursuing and using irrelevant information to make decisions they otherwise would not have made (84). Information presentation should be based on a scientific understanding of the impact of different formats on comprehension and decision making among those with differing levels of numeracy. Below we review two additional issues that should be considered when presenting numeric information to individuals with different degrees of numeracy, and suggest ways in which clinicians can present information in order to minimize the adverse effects of low numeracy.

Framing Effects

Although both low-numerate and high-numerate people are susceptible to framing effects, the less numerate appear to be more susceptible to these effects (10). “Framing” refers to how a decision problem is presented, and a framing effect occurs when a different but equivalent

description of the same decision problem yields different preferences. For example, describing an operation as having a 90% survival rate is equivalent to saying that the operation has a 10% mortality rate, yet surgery seems more attractive to decision makers when presented in terms of the gain frame of survival as opposed to the loss frame of mortality. This framing effect was demonstrated in a lung cancer treatment study in which both patients and physicians found surgery preferable to radiation therapy when the choice problem was framed as the probability of survival rather than the probability of dying (85). In another framing study (10), participants were asked to rate students based on their exam scores, which were described as either percent correct (the positive frame) or percent incorrect (the negative frame). As predicted, the framing effect was more pronounced among the less numerate participants, presumably because the more numerate participants were able to transform numbers from one format (36% incorrect) to the normatively equivalent frame (64% correct). Research has shown that framing effects remain when numbers are eliminated (e.g., replacing “600 people die” with a verbal description, such as “some people die”) (69,86). Thus, it appears to be the qualitative interpretation of numbers, rather than the numbers themselves, that determines choices. (Qualitative *interpretations* of numbers are the default mode of processing, and are not the same as qualitative *expressions* of numbers using phrases, such as “often” or “infrequent.”) Taken together, these studies suggest that highly numerate people are better able to extract meaningful qualitative relations from numerical information. The clinical implication of these findings is that clinicians should not assume that wording choices are benign, especially for low-numerate individuals. Although equivalent expressions of risk (e.g., 95% of patients survive surgery vs. 5% of patients die during surgery) should yield the same decision regardless of whether information is presented in the “survival” or “mortality” frame, research has shown that low- numerate individuals often make

different decisions depending upon how risk is framed. Therefore it may be prudent for clinicians to frame important medical information in different but equivalent ways.

Intertemporal Choice

Low-numerate and high-numerate individuals appear to access and use different sources of information when making decisions that have short-term and long-term consequences. People are often asked to incur concrete costs in the present (e.g., take a medication, experience anxiety about a medical test) in order to reap long-term - but abstract and probabilistic - rewards later. Research in economics suggests that compared to the more numerate, less numerate individuals tend to overweight short-term costs and benefits at the expense of long-term benefits (87,88). They also tend to base judgments and decisions on things they know for certain rather than on things that might happen in the future. Presumably this is because low-numerate individuals have a poor understanding of probabilistic information and are therefore less likely to attend to this kind of information. The implication of this for health is that low-numerate patients may tend to overweight events that occur in the near future and underweight or ignore events that might happen in the distant future. For example, consider the short-term costs and potential long-term rewards of statin medications used to lower cholesterol. There is the present cost and inconvenience of taking medicine on a daily basis versus the long-term benefit of lowered cholesterol, which, on an individual level, may or may not help to prevent a cardiovascular event. Because lowered cholesterol does not guarantee that a person will not experience a heart attack – it simply reduces one of many risk factors for heart attack – a low-numerate individual may not understand the value of taking medication for something that is not a “sure thing.”

Maintaining motivation and intention to adhere to medical advice in these kinds of situations may be difficult for low-numerate individuals.

THEORIES OF NUMERACY

Remarkable progress has been made in research on numeracy. Measures have been developed that correlate with medical outcomes and with psychological phenomena such as framing effects and temporal discounting, providing insights into how numeracy affects processing of quantitative information. Further insights are provided by identifying key mediators, such as math phobia, that are known to compromise quantitative performance (89), and other key factors that are correlated with numeracy, such as literacy, education, age (older people are more likely to be low in numeracy), and race/ethnicity (Hispanics tend to be lower in numeracy) (23). However, much of this research is descriptive and unmotivated by theory. In other words, these correlations do not identify causal mechanisms that lead to specific predictions other than “predicting” what has previously been found. For example, although researchers have identified graphic formats that seem to reduce certain types of errors, these effects have been obtained largely through trial and error, rather than being motivated by theory.

Nevertheless, theoretical frameworks exist that could be used to generate predictions about numeracy, and recent research has begun to exploit this possibility (10). There are four major theoretical approaches that are relevant to numeracy: the traditional information-processing or computational approach that stresses precision, analysis, and elaboration to achieve accuracy (90); evolutionary approaches that stress natural quantitative processing as illustrated by frequency effects (91,92); dual-process approaches that contrast intuitive (or affective) and analytical processing in which errors are due mainly to intuitive processing; and fuzzy-trace

theory, another dual-process theory, but one that stresses gist-based intuition as an advanced mode of processing and contrasts it with verbatim-based analytical processing (69). Space does not permit an exhaustive review of these theories regarding numeracy (but see 68). We can, however, briefly outline these approaches and their implications for numeracy, and point to avenues for future research.

According to the information-processing approach, working memory limitations interfere with numerical (and other) processing. Therefore, efforts to reduce the burden on working memory are predicted to improve performance. In this view, health care providers can make numbers more understandable and useable to patients by reducing the cognitive effort needed to process information - for example, by reducing the impact of irrelevant sources of information on judgments (83). Also, processing information more precisely and more elaborately ought to improve performance because number “crunching” or computation is the ideal. Consistent with this approach, asking people to actively process information by enumerating reasons for their preferences or by indicating the size of a risk on a bar chart may enhance the use and comprehension of numbers and reduce reliance on other sources of information (e.g., anecdotes and testimonials) (94,95). Further, Mazzocco et al. (94) found that asking decision makers to state reasons for their choices encouraged greater weighting of numerical information over less relevant, non-numerical sources of information, such as emotion and anecdotes. Many public health programs emphasize this kind of precise and elaborate processing of numerical information. Modern dual-process theories, discussed below, have subsumed this computational approach into their assumptions about the analytical side of processing (70).

Evolutionary approaches suggest that there are two modes of quantitative processing: one that is evolutionarily ancient and involves frequencies (91) or crude quantitative distinctions

(92), and another that is more recent and involves formal, numerical knowledge (e.g., knowledge of probabilities). Some promising work is being conducted comparing numerical processing in primates and young children to that of adult humans (96), but this research has not developed to the point where specific predictions can be made about numeracy. Research on so-called natural frequencies, although once quite promising, has not panned out empirically; a growing body of evidence challenges the assumption that frequencies are any easier to understand than probabilities (for reviews, see 68,97). Indeed, the latter claim has been tested repeatedly in the context of health decision making and been disconfirmed. Studies that initially reported a difference between frequencies and probabilities have been criticized on methodological grounds (68,97).

Most modern theorists take a dual-process approach in order to account for conflicting patterns of performance in reasoning and decision making (98). Epstein (99) developed a series of measures of intuitive or “experiential” thinking versus analytical or rational thinking, which have been extended by Slovic and colleagues (100). The basic assumption of this approach is that intuitive thinking is the source of biases and errors in numerical processing, accounting for framing effects and ratio biases. Analytical thinking is believed to be the source of accurate and objective numerical processing. (These dual-process claims about accuracy are central in the articles we discuss, although it should be acknowledged that intuition in these approaches is not assumed to inevitably lead to errors.) Affect, defined as either mood or valence (i.e., good-bad), is thought to be an aspect of intuition. According to Peters and colleagues, high-numerate and low-numerate people appear to access different sources of information when making decisions. The low numerate, who are less likely to attend to and understand numbers, appear to be informed more by non-numeric, sometimes extraneous, sources of information, such as mood or

affect. This effect was demonstrated in a study that examined how people made judgments about hospital quality. In this study, a community sample of 152 adults evaluated hypothetical hospital data. Low-numerate subjects were more likely to base their rating of hospital quality on their mood rather than numerical quality indicators, while high-numerate subjects showed no effect of mood; rather, the high numerate were able to use a numerical quality indicator to rate the hospitals (93). In another experiment, intuition was pitted against rational analysis by making an objectively worse choice more tempting (10). Subjects were offered a prize if they drew a colored jellybean from a bowl. Bowl A contained 9 colored and 91 white beans, and Bowl B contained 1 colored and 9 white beans. Consequently, the chance of picking a colored jellybean was objectively better if you picked from Bowl B (10% chance of winning) than if you picked from Bowl A (9% chance of winning). Despite this, 33% of low-numerate subjects and 5% of high-numerate subjects picked from Bowl A, which was clearly not the rational choice if one wanted to win a prize. High-numerate subjects tended to select Bowl B because they perceived the probability of winning more clearly than low-numerate subjects. The less numerate were influenced more by the number of *winning* beans than by the number of losing jelly beans (a ratio bias effect). Although dual-process or rational-affective theories motivated this experiment, it is unclear whether affect *per se* or alternative mechanisms involving class relations (102) is the source of these differences in processing (68). Nevertheless, these studies on mood and ratio bias suggest that compared to high-numerate patients, the preferences expressed by low-numerate patients are likely to be more labile and influenced by extraneous cues, and less influenced by objective probabilities.

Finally, fuzzy-trace theory was developed to explain challenges to several of the assumptions discussed above, notably that working memory capacity has been found to be

unrelated to reasoning performance in most paradigms (for a review, see 69). Fuzzy-trace theory explains framing effects (68,86) as a function of a fuzzy-processing preference; that is, people make decisions on the basis of bottom-line gist. Moreover, gist-based processing increases with experience and knowledge. Thus, framing effects were predicted and found to increase from childhood to adulthood (103); other heuristics and biases show a similar, counterintuitive trend (see Table 3, 104). In adulthood, experts have been found to base their decisions more on simple gist, compared to novices with less experience and knowledge (105).

Numerical processing in particular has been a longstanding focus of research on fuzzy-trace theory (76,102). For example, the ratio bias effect (preferring options with smaller probabilities but larger numerators) described earlier is an example of a class-inclusion illusion that involves retrieval and processing assumptions as well as representational assumptions (106). Any ratio concept, including probability, is inherently confusing because the referents of classes overlap. People focus on the target classes in numerators (e.g., the 9 colored jelly beans in Bowl A) and neglect the classes in the denominator (e.g., the 100 total jelly beans in Bowl A), producing the ratio bias effect (also called the numerosity effect) (76). Just as reminders about non-equivalence across frames can reduce framing effects, reminders about denominators can reduce the ratio bias effect. Class-inclusion errors persist for advanced reasoners. For example, physicians and high-school students performed equally poorly on a base-rate neglect problem, which is another type of class-inclusion problem. In this study, students and physicians were asked to estimate the probability of disease for a patient with a positive test result, given the base rate of disease and accuracy of the test (107). In sum, although traditional dual-process approaches generally equate accuracy with analysis, fuzzy-trace theory predicts that more experienced, knowledgeable reasoners (e.g., the more highly numerate) should rely mainly on

the simple, qualitative gist of numbers, but should not be less subject to class-inclusion illusions. Indeed, enumerating reasons for choices and focusing on fine-grained numerical distinctions as opposed to stark qualitative contrasts or gist (e.g., more patients die in Hospital A than Hospital B) leads to *lower* levels of performance in reasoning and decision making (69,105). Going with your gut - or gist - may be especially beneficial for older people whose verbatim memories are less robust: Age differences in choice quality between younger and older adults are reduced when decisions are based on gut feelings (108). Recently, findings on affect have been integrated with the principles of fuzzy-trace theory; such findings indicate that affect is frequently an essential part of the gist experience. Although extant research on numeracy is consistent with fuzzy-trace theory, research has not critically compared predictors of alternate dual-process approaches (109).

FUTURE RESEARCH IN HEALTH NUMERACY

We suggest that future research in health numeracy address conceptual, measurement, and communication issues. 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.

A theoretical framework is also needed in order to develop measures that fully operationalize the construct of health numeracy. Presently, the most widely used measures to assess health numeracy tap different dimensions of numeracy. For example, the TOFHLA, which is perhaps the most widely used of all the health numeracy instruments, does not test basic

arithmetic computation skills, but rather, simple, functional applied skills (such as understanding directions for taking medication). The TOFHLA differs considerably from the often cited Schwartz et al. general numeracy assessment (17) and the Lipkus et al. numeracy scale (21), both of which primarily test arithmetic computation skills, such as converting a percentage to a proportion and converting a proportion to a percentage. The Medical Data Interpretation Test (56), on the other hand, does not focus on arithmetic skills *per se*, but rather, on interpretation and deriving higher level inferences from numerical information. A novel approach to assessing health numeracy - the Numeracy Understanding in Medicine Instrument - is being developed by Schapira and colleagues (110). Based on an empirically derived health numeracy framework, Schapira is using Item Response Theory to develop a numeracy measure that will be cross-culturally equivalent across racial and ethnic groups. One of the advantages of this technique is that the final instrument will be appropriate for individuals at all skill levels, thereby making it potentially usable for screening (in the clinical setting) or research purposes. The instrument will also be able to identify an individual's strengths and weaknesses, which will in turn allow health care providers to tailor communication and intervention strategies. The need for a more comprehensive measure of health numeracy was underscored in a study by Donelle and colleagues (111), which elicited different levels of performance using different numeracy assessments.

It is also important to examine whether subjective numeracy measures, such as the SNS (60) and the STAT-Confidence scale (62) tap the same underlying dimensions as measures of objective numeracy. Limited and conflicting data make it difficult to evaluate how well subjective measures correlate with objective measures. For example, in one study involving 357 patients attending a university internal medicine clinic, 70% of subjects reported that they

considered themselves “good with numbers,” yet only 2% of subjects answered all three items on the Schwartz et al. objective numeracy measure correctly (18). In contrast, Fagerlin and colleagues reported that the SNS was significantly correlated with the Lipkus et al. objective numeracy scale ($r = 0.63 - 0.68$) (60). Clearly, further testing of subjective numeracy scales is needed. If these scales are found to be poor measures of numeracy, future research will need to develop and validate new subjective measures, or refine existing measures that could be used in the clinical setting. This would enable providers to screen for numeracy in a minimally burdensome way. Although screening for numeracy in a clinical setting is not currently practiced, health literacy screening in a primary care setting has been found to be acceptable to physicians (112).

Much work needs to be done in the area of communicating risk through visual and verbal means. As described in this paper, a number of researchers have examined the effects of graphic data displays on understanding health risks. Excellent reviews of this literature have been provided by Lipkus & Hollands (66) and Ancker and colleagues (65), and recommended practices have been proposed by Lipkus (113). Researchers in the health domain have studied how best to convey quantitative health risks by examining how manipulating graphic formats can improve accuracy, reduce response time, affect risk perception and behavior, and affect perceived credibility of the data (63,70,78,114-117). However, research in this area has been largely atheoretical (113). Future work should focus on developing theory based models of the perceptual and cognitive processes involved in understanding graphic displays of health risk. For example, such work in non-health domains (118,119) has been informed by theories from cognitive psychology, educational psychology, and gestalt psychology, but there has not been a concerted effort to coordinate and apply contemporary theories to the health domain. It will also

be important to examine the mechanisms by which level of numeracy interacts with graphic format to affect comprehension.

Clear communication between patients and physicians, and among medical professionals, regarding the potential risks and benefits of treatment is essential for informed decision making. However, whether it is most effective to communicate risk using numbers (e.g., a 0.001% chance of experiencing a side effect), qualitative expressions of likelihood (e.g., a rare side effect), or both modes is an important research question (120). It has been suggested that qualitative expressions of probability may be preferred because they seem to be natural and easy to use, while quantitative expressions of probability may be preferred because they are more precise (121). A number of studies have demonstrated that there is considerable variation in how patients and physicians interpret verbal probability expressions (122-124). Patients and the lay public differ as to their preference for quantitative or qualitative risk communication (125-128). In one study, physicians preferred to use verbal expressions of probability, while laypersons found numbers more helpful (129). Some researchers have suggested that qualitative expressions of probability could be systematically codified for use in the clinical encounter (130,131), while others are emphatic that they should be eliminated entirely in the medical context (132). Mapes (133) observed that physicians' interpretations of probability expressions depended on the therapeutic context. In his study, a "rare" side effect was judged to be much more severe when it was associated with a beta-blocker than when it was associated with an antihistamine. Whether one mode of communicating risk leads to more effective communication remains an empirical question. In addition, it will be important to examine how interpretation of verbal and numeric expressions of probability is influenced by level of numeracy, context, and experience (128).

CONCLUSIONS

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. Graphic displays are of general benefit, depending on the relation to be conveyed (e.g., lines to convey trends, heights of bar graphs to convey magnitude), and these visual displays may be especially helpful to those who are least numerate. Making numbers more transparent should offset the use of irrelevant cues, such as mood, in medical decision making, which are relied on more by the low numerate. The low numerate are also more subject to sharper temporal discounting - weighing immediate rewards much more than temporally distant rewards - presumably because they are less able to understand probabilistic risk. Because prevention of the leading causes of death (e.g., cancer and cardiovascular disease) depends in large part on taking action now to prevent serious consequences later, the low-numerate patient is likely to require more extensive explanations of risk in order to engage in prevention behaviors. Although the highly numerate have been less subject to some biases (such as preferring lower probability options with larger, enticing numerators), they may achieve this superior numerical processing by, ironically, focusing on qualitative gist (rather than quantitative details). The role of affect in guiding intuitions, even among the highly numerate, is a promising avenue for future research, and one

that clinicians may someday use to empower patients to take advantage of the evidence in evidence-based medicine.

REFERENCES

- (1) Health Information National Trends Survey (HINTS 2005). National Cancer Institute. Retrieved December 10, 2007, from <http://hints.matthewsgroup.com/register/register.aspx>.
- (2) Kirsch I, Jungeblut A, Jenkins L, Kolstad A: Adult literacy in America: A first look at the findings of the National Adult Literacy Survey. Washington, DC: National Center for Education Statistics, U.S. Department of Education, 1993.
- (3) Kutner M, Greenberg E, Jin Y, Paulsen C: The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). Washington, DC: National Center for Education Statistics, U.S. Department of Education, 2006.
- (4) U.S. Department of Health and Human Services: Healthy People 2010. 2nd ed. With understanding and improving health and objectives for improving health. 2 vols. Washington, DC: U.S. Government Printing Office, November 2000.
- (5) Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association: Health literacy: Report of the Council on Scientific Affairs. Journal of the American Medical Association. 1999, 281:552-557.
- (6) Nutbeam D: Health promotion glossary. Health Promotion International. 1998, 13:349-364.

- (7) Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke SE: A definition and operational framework for health numeracy. American Journal of Preventive Medicine. 2005, 29:375-376.
- (8) Ancker JS, Kaufman D: Rethinking health numeracy: A multidisciplinary literature review. Journal of the American Medical Informatics Association. 2007, 14:713-721.
- (9) Reyna VF, Brainerd CJ: The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. Learning and Individual Differences. 2007, 17:147-159.
- (10) Peters E, Vastfjall D, Slovic P, et al.: Numeracy and decision making. Psychological Science. 2006, 17:408-414.
- (11) Hibbard JH, Peters E, Dixon A, Tusler M: Consumer competencies and the use of comparative quality information: It isn't just about literacy. Medical Care Research and Review. 2007, 64:379-394.
- (12) Kutner M, Greenberg E, Baer J: A First Look at the Literacy of America's Adults in the 21st Century (NCES 2006-470). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2005.
- (13) Williams MV, Parker RM, Baker DW, et al.: Inadequate functional health literacy among patients at two public hospitals. Journal of the American Medical Association. 1995, 274:1677-1682.

- (14) Rothman RL, Housam R, Weiss H, et al: Patient understanding of food labels: The role of literacy and numeracy. 2006. American Journal of Preventive Medicine. 2006, 31:391-398.
- (15) Davis TC, Crouch MA, Long SW, et al.: Rapid assessment of literacy levels of adult primary care patients. Family Medicine. 1991, 23:433-435.
- (16) Wilkinson G: WRAT-3: Wide Range Achievement Test, Administration Manual. Wilmington, DE: Wide Range, Inc., 1993.
- (17) Schwartz LM, Woloshin S, Black WC, Welch HG: The role of numeracy in understanding the benefit of screening mammography. Annals of Internal Medicine. 1997, 127:966-972.
- (18) Sheridan SL, Pignone MP, Lewis CL: A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. Journal of General Internal Medicine. 2003, 18:884-892.
- (19) Forrow L, Taylor WC, Arnold RM: Absolutely relative: How research results are summarized can affect treatment decisions. American Journal of Medicine. 1992, 92:121-124.
- (20) Sheridan SL, Pignone M: Numeracy and the medical students' ability to interpret data. Effective Clinical Practice. 2002, 5:35-40.
- (21) Lipkus IM, Samsa G, Rimer BK: General performance on a numeracy scale among highly educated samples. Medical Decision Making. 2001, 21:37-44.

- (22) Yamigishi K: When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. Applied Cognitive Psychology. 1997, 11:495-506.
- (23) Reyna V F, Lloyd F J, Whalen P. Genetic testing and medical decision making. Archives of Internal Medicine. 2001, 161:2406-2408.
- (24) Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM: The framing effect of relative and absolute risk. Journal of General Internal Medicine. 1993, 8:543-548.
- (25) Chao C, Studts JL, Abell T, et al.: Adjuvant chemotherapy for breast cancer: How presentation of recurrence risk influences decision making. Journal of Clinical Oncology. 2003, 21:4299-4305.
- (26) Naylor CD, Chen E, Strauss B: Measured enthusiasm: Does the method of reporting trial results alter perceptions of therapeutic effectiveness? Annals of Internal Medicine. 1992, 117:916-921.
- (27) Edwards A, Elwyn G, Covey J, Matthews E, Pill R: Presenting risk information: A review of the effects of “framing” and other manipulations on patient outcomes. Journal of Health Communication. 2001, 6:61-682.
- (28) Bucher HC, Weinbacher M, Gyr K: Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. British Medical Journal. 1994, 309:761-764.
- (29) Hux JE, Naylor CD: Communicating the benefits of chronic preventive therapy: Does the format of efficacy data determine patients’ acceptance of treatment? Medical Decision Making. 1995, 15:152-157.

- (30) Sarfati D, Howden-Chapman P, Woodward A, Salmond C: Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. Journal of Medical Screening. 1998, 5:137-140.
- (31) DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP: Literacy and health outcomes: A systematic review of the literature. Journal of General Internal Medicine. 2004, 19:1228-1239.
- (32) Berkman ND, DeWalt DA, Pignone MP, et al.: Literacy and Health Outcomes. Summary, Evidence Report/Technology Assessment No. 87 (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-1. Rockville, MD: Agency for Healthcare Research and Quality, January 2004.
- (33) Baker DW, Parker RM, Williams MV, Clark WS, Nurss J: The relationship of patient reading ability to self-reported health and use of health services. American Journal of Public Health. 1997, 87:1027-1030.
- (34) Sentell TL, Halpin HA: Importance of adult literacy in understanding health disparities. Journal of General Internal Medicine. 2006, 21:862-866.
- (35) Williams MV, Baker DW, Parker RM, Nurss JR: Relationship of functional health literacy to patients' knowledge of their chronic disease. Archives of Internal Medicine. 1998, 158:166-172.

- (36) Parker RM, Baker DW, Williams MV, Nurss JR: The test of functional health literacy in adults. Journal of General Internal Medicine. 1995, 10:537-541.
- (37) Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J: Development of a brief test to measure functional health literacy. Patient Education and Counseling. 1999, 38:33-42.
- (38) Baker DW, Parker RM, Williams MV, Clark WS: Health literacy and the risk of hospital admission. Journal of General Internal Medicine. 1998, 13:791-798.
- (39) Scott TL, Gazmararian JA, Williams MV, Baker DW: Health literacy and preventive health care use among Medicare enrollees in a managed care organization. Medical Care. 2002, 40:395-404.
- (40) Estrada CA, Martin-Hryniwicz M, Peek BT, Collins C, Byrd JC: Literacy and numeracy skills and anticoagulation control. American Journal of Medical Science. 2004, 328:88-93.
- (41) Apter AJ, Cheng J, Small D, et al.: Asthma numeracy skill and health literacy. Journal of Asthma. 2006, 43:705-710.
- (42) Gazmararian JA, Kripalani S, Miller MJ, et al.: Factors associated with medication refill adherence in cardiovascular-related diseases. Journal of General Internal Medicine. 2006, 21:1215-1221.
- (43) Lenert LA, Sherbourne CD, Reyna VF: Utility elicitation using single item questions compared with a computerized interview. Medical Decision Making. 2001, 21:97-104.

- (44) Morimoto T, Fukui T: Utilities measured by rating scale, time trade-off, and standard gamble: Review and reference for health care professionals. Journal of Epidemiology. 2002, 12:160-178.
- (45) Woloshin S, Schwartz LM, Moncur M, Gabriel S, Tosteson ANA: Assessing values for health: Numeracy matters. Medical Decision Making. 2001, 21:382-390.
- (46) Schwartz SR, McDowell J, Yueh B: Numeracy and the shortcomings of utility assessment in head and neck cancer patients. Head and Neck. 2004, 26:401-407.
- (47) Lindau ST, Tomori C, Lyons T, et al.: The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. American Journal of Obstetrics and Gynecology. 2002, 186:938-943.
- (48) Bass PF, Wilson JF, Griffith CH, Barnett DR: Residents' ability to identify patients with poor literacy skills. Academic Medicine. 2002, 77:1039-1041.
- (49) Rogers ES, Wallace LS, Weiss BD: Misperceptions of medical understanding in low-literacy patients: Implications for cancer prevention. Cancer Control. 2006, 3:225-229.
- (50) Doak LG, Doak CC: Patient comprehension profiles: Recent findings and strategies. Patient Counseling and Health Education. 1980, 2:101-106.
- (51) Kicklighter JR, Stein MA: Factors influencing diabetic clients' ability to read and comprehend printed diabetic diet material. The Diabetes Educator.

- 1993, 19:40-46.
- (52) Davis TC, Crouch MA, Wills G, Miller S, Abdehou DM: The gap between patient reading comprehension and the readability of patient education materials. The Journal of Family Practice. 1990, 31:533-538.
- (53) Powers RD: Emergency department patient literacy and the readability of patient-directed materials. Annals of Emergency Medicine. 1988, 17:124-126.
- (54) Safer RS, Keenan J: Health literacy: The gap between physicians and patients. American Family Physician. 2005, 72:463-468.
- (55) Dunning D, Heath C, Suls JM: Flawed self-assessment: Implications for health, education, and the workplace. Psychological Science in the Public Interest. 2004, 5:69-106.
- (56) Schwartz LM, Woloshin S, Welch HG: Can patients interpret health information? An assessment of the Medical Data Interpretation Test. Medical Decision Making. 2005, 25:290-300.
- (57) Weiss BD, Mays MZ, Martz W, et al.: Quick assessment of literacy primary care: The Newest Vital Sign. Annals of Family Medicine. 2005, 3:514-522.
- (58) Estrada C, Barnes V, Collins C, Byrd JC: Health literacy and numeracy. Journal of the American Medical Association. 1999, 282:527.
- (59) Montori VM, Leung TW, Thompson CA, et al.: Validation of a diabetes numeracy evaluation tool [Abstract]. Diabetes. 2004, 53(Suppl 2):A224-A225.

- (60) Fagerlin A, Zikmund-Fisher BJ, Ubel PA, et al.: Measuring numeracy without a math test: Development of the subjective numeracy scale. Medical Decision Making. 2007, 27:672-680.
- (61) Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A: Validation of the Subjective Numeracy Scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. Medical Decision Making. 2007, 27:663-671.
- (62) Woloshin S, Schwartz LM, Welch HG: Patients and medical statistics: Interest, confidence, and ability. Journal of General Internal Medicine. 2005, 20:996-1000.
- (63) Feldman-Stewart D, Kocovski N, McConnell BA, et al: Perception of quantitative information for treatment decisions. Medical Decision Making. 2000, 20:228-238.
- (64) Mazur DJ, Merz JF: How the manner of presentation of data influences older patients in determining their treatment preferences. Journal of the American Geriatric Society. 1993, 41:223-228.
- (65) Ancker J, Senathirajah Y, Kukafka R, Starren J: Design features of graphs in health risk communication: A systematic review. Journal of the American Medical Informatics Association. 2006, 13:608-618.
- (66) Lipkus IM, Hollands JG: The visual communication of risk. Journal of the National Cancer Institute Monograph. 1999, 25:149-163.

- (67) Cleveland W, McGill R: Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association. 1984, 77:541-547.
- (68) Reyna VF, Brainerd CJ: Numeracy, ratio bias and denominator neglect in judgments of risk and probability. Learning and Individual Differences (in press, 2008).
- (69) Reyna V, Brainerd C: Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences. 1995, 7:1-75.
- (70) Reyna VF: A theory of medical decision making and health: Fuzzy-trace theory. Medical Decision Making (in press, 2008).
- (71) Feldman-Stewart D, Brundage MD, Zotov V: Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. Medical Decision Making. 2007, 27:34-43.
- (72) Hollands JG, Tanaka T, Dyre BP: Understanding bias in proportion production. Journal of Experimental Psychology: Human Perception and Performance. 2002, 28:563-574.
- (73) Sandman P, Weinstein N, Miller P: High risk or low: How location on a “risk ladder” affected perceived risk. Risk Analysis. 1994, 14:35-45.
- (74) Lakoff G: Women, Fire, and Dangerous Things. Chicago, IL: University of Chicago Press, 1987.
- (75) Zikmund-Fisher BJ, Fagerlin A, Ubel PA: What’s time got to do with it? Inattention to duration in interpretation of survival graphs. Risk Analysis. 2005, 25:1-7.

- (76) Reyna VF, Brainerd CJ: The origins of probability judgment: A review of data and theories. In: Wright G, Ayton P (eds), Subjective Probability. New York, NY: Wiley, 1994, 239-272.
- (77) Stone E, Yates J, Parker A: Effects of numerical and graphical displays on professed risk-taking behavior. Journal of Experimental Psychology Applied. 1997, 3:243-256.
- (78) Stone E, Sieck W, Bull B, et al.: Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. Organizational Behavior and Human Decision Processes. 2003, 90:19-36.
- (79) Schapira MM, Nattinger AB, McHorney CA: Frequency or probability? A qualitative study of risk communication formats used in health care. Medical Decision Making. 2001, 21:459-467.
- (80) Zacks J, Levy E, Tversky B, Schiano D: Reading bar graphs: Effects of extraneous depth cues and graphical context. Journal of Experimental Psychology Applied. 1998, 4:119-138.
- (81) Levy E, Zacks J, Tversky B, Schiano D: Gratuitous graphics? Putting preferences in perspective. Conference on Human factors in computing systems. Vancouver, BC, 1996, 42-49.
- (82) Fagerlin A, Zikmund-Fisher B, Ubel P, Smith D: Measuring numeracy and the impact of numeracy on medical decision making. Annual Meeting of the Society of Behavioral Medicine. Washington, DC: 2007.

- (83) Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK: Less is more in presenting quality information to consumers. Medical Care Research and Review. 2007, 64:169-190.
- (84) Bastardi A, Shafir E: On the pursuit and misuse of useless information. Journal of Personality and Social Psychology. 1998, 75:19-32.
- (85) McNeil BJ, Pauker SG, Sox HC, Tversky A: On the elicitation of preferences for alternative therapies. New England Journal of Medicine. 1982, 306:1259-1262.
- (86) Reyna VF, Brainerd CJ: Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. Journal of Behavior and Decision Making. 1991, 4:249-262.
- (87) Benjamin DJ, Brown SA, Shapiro JM: "Who is "behavioral"? Cognitive ability and anomalous preferences. Retrieved December 17, 2007, from SSRN: <http://ssrn.com/abstract=675264>.
- (88) Frederick S: Cognitive reflection and decision making. Journal of Economic Perspectives. 2005, 19: 24-42.
- (89) Ashcraft MH: Math anxiety: Personal, educational, and cognitive consequences. Current Directions in Psychological Science. 2002, 11:181-185.
- (90) Fischhoff B: Assessing adolescent decision-making competence. Developmental Review (in press, 2008).
- (91) Gigerenzer G: Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In Wright G, Ayton P (eds). Subjective Probability. New York, NY: Wiley, 1994, 129-161.

- (92) Dehaene S: The Number Sense: How the Mind Creates Mathematics. Oxford: Oxford University Press, 1997.
- (93) Peters E, Dieckmann N, Vastfjall D, et al.: Bringing meaning to numbers: The function of affect in choice. Unpublished manuscript.
- (94) Mazzocco K, Peters E., Bonini N., Slovic, P, Cherubini P: How priming analytical versus emotional thinking influences choice processes. Unpublished manuscript.
- (95) Natter HM, Berry DC: Effects of active information processing on the understanding of risk information. Applied Cognitive Psychology. 2005, 19:123-135.
- (96) Chen MK, Lakshminaryanan V, Santos LR. The evolution of our preferences: Evidence from capuchin monkey trading behavior. Journal of Political Economy. 2006, 114:517-537.
- (97) Barbey AK, Sloman SA. Base-rate respect: From ecological rationality to dual processes. Behavioral and Brain Sciences. 2007, 30:241-254.
- (98) Sloman S: The empirical case for two systems of reasoning. Psychological Bulletin. 1996, 119:3-22.
- (99) Epstein S. Integration of the cognitive and psychodynamic unconscious. American Psychologist. 1994, 49:709-724.
- (100) Slovic P, Finucane ML, Peters E, MacGregor DG: Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. Risk Analysis. 2004, 24:311-322.

- (101) Peters E: Psychological mechanisms of numeracy decisions. Annual Meeting of the Society of Behavioral Medicine. Washington, DC: 2007.
- (102) Reyna VF, Mills BA: Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). Behavioral and Brain Sciences. 2007, 30:278-280.
- (103) Reyna VF, Ellis SC: Fuzzy-trace theory and framing effects in children's risky decision making. Psychological Science. 1994, 5:275-279.
- (104) Reyna VF, Farley F: Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. Psychological Science in the Public Interest. 2006, 7:1-44.
- (105) Reyna VF, Lloyd FJ: Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. Journal of Experimental Psychology: Applied. 2006, 12:179-195.
- (106) Reyna VF: Class inclusion, the conjunction fallacy, and other cognitive illusions. Developmental Review. 1991, 11:317-336.
- (107) Reyna VF, Adam MB: Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. Risk Analysis. 2003, 23:325-342
- (108) Mikels JA, Loeckenhoff C, Maglio S, et al.: Going with your gut may pay off for older people: Age differences in choice quality are reduced when decisions are based on feelings. Unpublished manuscript.
- (109) Rivers S, Mills B, Reyna VF: Risky behavior under the influence: A fuzzy-trace theory of emotion in adolescence. Developmental Review (in press, 2008).

- (110) Schapira M: The development of a health numeracy measure. Computer Retrieval of Information on Scientific Projects. Retrieved December 13, 2007, from <http://crisp.cit.nih.gov>.
- (111) Donelle L, Hoffman-Goetz L, Arocha JF: Assessing health numeracy among community-dwelling older adults. Journal of Health Communication. 2007, 12:651-655.
- (112) Seligman HK, Wang FF, Palacios JL, et al.: Physician notification of their diabetes patients' limited health literacy: A randomized controlled trial. Journal of General Internal Medicine. 2005, 20:1001-1007.
- (113) Lipkus IM: Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. Medical Decision Making. 2007, 27:696-713.
- (114) Feldman-Stewart D, Brundage MD, Zotov V: Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. Medical Decision Making. 2007, 27:34-43.
- (115) Zikmund-Fisher BJ, Fagerlin A, Ubel PA: Mortality versus survival graphs: Improving temporal consistence in perceptions of treatment effectiveness. Patient Education and Counseling. 2007, 66:100-107.
- (116) Schapira MM, Nattinger AB, McAuliffe TL: The influence of graphic format on breast cancer risk communication. Journal of Health Communication. 2006, 11:569-582.

- (117) Stone ER, Yates JF, Parker AM: Risk communication: Absolute versus relative expressions of low-probability risks. Organizational Behavior and Human Decision Processes. 1994, 60:387-408.
- (118) Carpenter PA, Shah P: A model of the perceptual and conceptual processes in graph comprehension. Journal of Experimental Psychology: Applied. 1998, 4:75-100.
- (119) Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. Science. 1985, 229:828-833.
- (120) Burkell J: What are the chances? Evaluating risk and benefit information in consumer health materials. Journal of the Medical Library Association. 2004, 92:200-208.
- (121) Wallsten TS, Budescu DV, Zwick R, Kemp SM: Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bulletin of the Psychonomic Society. 1993, 31:135-138.
- (122) Mazur DJ, Hickam DH: Patient interpretations of terms connoting low probabilities when communicating about surgical risk. Theoretical Surgery. 1993, 8:143-145.
- (123) Kenney RM: Between never and always. New England Journal of Medicine. 1981, 305:1097-1098.
- (124) Bryant GD, Norman GR: Expressions of probability: Words and numbers. New England Journal of Medicine. 1980, 302:411-412.
- (125) Woloshin KK, Ruffin MT, Gorenflo DW: Patients' interpretation of qualitative probability statements. Archives of Family Medicine. 1994, 3:961-966.

- (126) Mazur DJ, Hickam DH, Mazur MD: How patients' preferences for risk information influence treatment choice in a case of high risk and high therapeutic uncertainty: Asymptomatic localized prostate cancer. Medical Decision Making. 1999, 19:394-398.
- (127) Mazur DJ, Hickam DH. Patients' preferences for risk disclosure and role in decision making for invasive medical procedures. Journal of General Internal Medicine. 1997, 12:114-117.
- (128) Mazur DJ, Hickam DH: Patients' interpretations of probability terms. Journal of General Internal Medicine. 1991, 6:237-240.
- (129) Brun W, Teigen KH: Verbal probabilities: Ambiguous, context-dependent, or both? Organizational Behavior and Human Decision Processes. 1988, 41:390-404.
- (130) O'Brien BJ: Words or numbers? The evaluation of probability expressions in general practice. Journal of the Royal College of General Practitioners. 1989, 39:98-100.
- (131) Kong A, Barnett GO, Mosteller F, Youtz C: How medical professionals evaluate expressions of probability. New England Journal of Medicine. 1986, 315:740-744.
- (132) Nakao MA, Axelrod S: Numbers are better than words: Verbal specifications of frequency have no place in medicine. The American Journal of Medicine. 1983, 74:1061-1065.
- (133) Mapes REA: Verbal and numerical estimates of probability in therapeutic contexts. Social Science and Medicine. 1979, 13A:277-282.