Published in final edited form as:
Psychol Bull. 2009 November ; 135(6): 943–973. doi:10.1037/a0017327.

How Numeracy Influences Risk Comprehension and Medical Decision Making

Valerie F. Reyna,
Departments of Human Development and Psychology, Cornell University;

Wendy L. Nelson,
National Cancer Institute, Bethesda, Maryland;

Paul K. Han, and
National Cancer Institute, Bethesda, Maryland;

Nathan F. Dieckmann
Decision Research, Eugene, Oregon, and University of Oregon

Abstract

We review the growing literature on health numeracy, the ability to understand and use numerical information, and its relation to cognition, health behaviors, and medical outcomes. Despite the surfeit of health information from commercial and noncommercial sources, national and international surveys show that many people lack basic numerical skills that are essential to maintain their health and make informed medical decisions. Low numeracy distorts perceptions of risks and benefits of screening, reduces medication compliance, impedes access to treatments, impairs risk communication (limiting prevention efforts among the most vulnerable), and, based on the scant research conducted on outcomes, appears to adversely affect medical outcomes. Low numeracy is also associated with greater susceptibility to extraneous factors (i.e., factors that do not change the objective numerical information). That is, low numeracy increases susceptibility to effects of mood or how information is presented (e.g., as frequencies vs. percentages) and to biases in judgment and decision making (e.g., framing and ratio bias effects). Much of this research is not grounded in empirically supported theories of numeracy or mathematical cognition, which are crucial for designing evidence-based policies and interventions that are effective in reducing risk and improving medical decision making. To address this gap, we outline four theoretical approaches (psychophysical, computational, standard dual-process, and fuzzy trace theory), review their implications for numeracy, and point to avenues for future research.

Keywords

risk perception; risk communication; mathematical cognition; intuition; dual processes

In a series of television and print advertisements, Robert Jarvik, inventor of the artificial heart, described the benefits of Lipitor for cardiovascular health. In one 2007 advertisement, Jarvik stands in front of an image of a heart. Next to him, in large print, the copy reads: “In patients with multiple risk factors for heart disease, Lipitor reduces risk of heart attack by 36%.*” If you failed to pay attention to the asterisk, you would have missed the following explanation for the impressive 36%: “*That means in a

Correspondence concerning this article should be addressed to Valerie F. Reyna, Department of Human Development, Cornell University, B44 Martha Van Rensselaer Hall, Ithaca, NY 14853. vr53@cornell.edu.

large clinical study, 3% of patients taking a sugar pill or placebo had a heart attack compared to 2% of patients taking Lipitor."

People have unprecedented access to information—available online, in print, and through other media—that they can use to improve their mental and physical health. Much of that information is expressed numerically. For example, the effectiveness of cancer treatments is expressed as survival rates (e.g., the percentage of treated patients who survive for 5 years), the benefits of lifestyle changes as reductions in cardiovascular risk, and the side effects of medications as probabilities of death, discomfort, and disability (Baker, 2006; Woloshin, Schwartz, & Welch, 2005). Indeed, numerical information about health is almost impossible to avoid, ranging from the cereal box at breakfast touting a four-point reduction in total cholesterol to direct-to-consumer advertisements in magazines reporting a 36% reduction in the risk of heart attack in the latest study of a cholesterol-lowering drug. The ubiquity and complexity of health-related numerical information place demands on people that, our review suggests, they are ill-prepared to meet.

Two recent trends in health care have exacerbated these demands. First, medical decision making has shifted from a mainly provider-centered to a shared or patient-centered model (e.g., Apter et al., 2008; Sheridan, Harris, & Woolf, 2004). Thus, there is an increased burden on patients to understand health-related information in order to make fully informed choices about their medical care. Second, there is an increased emphasis on applying research findings to achieve evidence-based health practices (Nelson, Reyna, Fagerlin, Lipkus, & Peters, 2008). Thus, people are routinely exposed to research findings with health implications, and health care providers must effectively convey these research findings to patients, findings that are often described numerically (Reyna & Brainerd, 2007). Unfortunately, numerical information is a particularly difficult form of information for both patients and health care providers to understand. As this review shows, low numeracy is pervasive and constrains informed patient choice, reduces medication compliance, impedes access to treatments, impairs risk communication (limiting prevention efforts among those most vulnerable to health problems), and, based on the scant research conducted on outcomes, appears to adversely affect medical outcomes.

To minimize the damaging effects of low numeracy, research on how people process numerical information, and how such processing can be improved, is essential. These questions—about how information is processed and can be improved—are fundamentally causal. However, most work on health numeracy has been descriptive rather than concerned with causal mechanisms, and we therefore lack sufficient understanding of how to improve numeracy in people facing various medical decisions. Thus, to resolve the dilemma of health numeracy—that people are swamped with numerical information that they do not understand, and yet they have to make life-and-death decisions that depend on understanding it—theory-driven research that tests causal hypotheses is of the first importance. Therefore, a major goal of this review is to spur interest in conducting such research.

Scope and Goals of the Review

Systematic research on numeracy has been growing steadily over the last several years, but there has not been a comprehensive published review of this literature. In addition to summarizing key findings, this review identifies gaps in our knowledge and suggests paths for future research in the field. The primary goal of this article is to review current directions in numeracy research and, in particular, to examine the relationship between numeracy and decision making in health and selected nonhealth domains with a view to establishing a foundation for future research on causal mechanisms.

In the first section of this review, we detail specific conceptualizations of numeracy that are referred to in the remainder of the article. Then, we consider the measurement of numeracy and the implications of different assessments for different conceptualizations of numeracy. In the next section, we describe national assessments of numeracy: how numeracy stacks up against other essential information-processing skills such as prose literacy; how numeracy differs in vulnerable subgroups of the population, such as the old and the poor; and how aspects of numeracy, such as understanding fractions, pose special challenges.

In the latter sections of the article, we discuss instruments that assess numeracy in individuals or samples of research subjects, as opposed to national surveys; these assessments also reveal low levels of understanding. We discuss how these assessments relate to risk perception, patient values for health outcomes, other judgments and decision making, health behaviors, and, finally, medical outcomes. Then we review selected research from the cognitive and developmental literatures that elucidates psychological mechanisms in numeracy as well as theories of mathematical cognition that bear on judgment and decision making, including affective approaches, fuzzy trace theory, and other dual-process perspectives, and evolutionary and neuroscience frameworks. Last, we summarize the current state of knowledge concerning numeracy and discuss possible future directions for the field.

Method

Literature Search

Several methods were used to search the literature for potentially relevant research reports. Electronic databases (e.g., PsycINFO, Medline) were used to capture an initial set of potentially relevant research reports. The initial search terms were relatively broad (numeracy, numerical ability, number ability, etc.), resulting in a large number of potential reports. We scanned the abstracts of all the articles identified in the electronic databases for inclusion in the review. After the initial search, we used the Web of Science database to identify additional reports that had referenced many of the pivotal numeracy articles. Finally, the reference lists of all articles identified by the first two methods were examined for additional articles that were missed by the electronic searches.

Study Inclusion

We focused primarily on empirical reports published in peer-reviewed journals or published books. We also excluded articles that reported single-case studies, introspective studies, and articles with very small sample sizes (e.g., results from interviews with two or three participants). A few unpublished working articles or other reports were included, but we did not make a specific effort to retrieve unpublished literature. We think the decision not to specifically seek unpublished reports is justified, as the primary purpose of this review was to get a broad sense of our current knowledge concerning numeracy and to propose directions for further research. This decision avoids such problems as overinterpretation of null effects (failures to detect effects that can be due to inadequate measures and methods), but it does leave open problems of publication bias (also called the “file-drawer problem”; Rosenthal, 1979).

Numeracy: Background

Increasing amounts of health information are being made available to the public, with the expectation that people can use it to reduce their risks and make better medical decisions. For example, patients are expected to take advantage of information about drug options available through Medicare Part D, assess the benefits and drawbacks of each option, and ultimately make wise choices regarding their care (Reed, Mikels, & Simon, 2008). The torrent of health

information is likely to persist because it is generated by multiple trends, such as the public's increasing demand for health information related to preventing diseases and making medical decisions; ongoing efforts of government agencies to create and disseminate health information; the proliferation of technologies that support rapid dissemination of research discoveries; and continuing efforts of the health care industry to promote adoption of various medical interventions, exemplified in direct-to-consumer advertising (e.g., Hibbard, Slovic, Peters, Finucane, & Tusler, 2001; Reyna & Brainerd, 2007; Woloshin, Schwartz, & Welch, 2004). Rising health care costs have also encouraged a more consumer-driven approach to health care, in which patients share in both decision making and associated costs, adding to the need for health information (Hibbard & Peters, 2003; but see Shuchman, 2007).

Researchers have long recognized the importance of literacy for making informed health decisions (Rudd, Colton, & Schacht, 2000). Individuals with limited literacy skills are at a marked disadvantage in this information age. Low literacy is associated with inferior health knowledge and disease self-management skills, and worse health outcomes (Baker, Parker, Williams, & Clark, 1998; Baker, Parker, Williams, Clark, & Nurss, 1997; Gazmararian, Williams, Peel, & Baker, 2003; Schillinger et al., 2002; Wolf, Gazmararian, & Baker, 2005).

A basic understanding of numerical concepts is arguably as important for informed decision making as literacy. In addition to basic reading and writing skills, people need an understanding of numbers and basic mathematical skills to use numerical information presented in text, tables, or charts. However, numeracy, the ability to understand and use numbers, has not received the same attention as literacy in the research literature. We describe national results in detail in a subsequent section, but it is instructive to note here that simple skills cannot be taken for granted. National surveys indicate that about half the U.S. population has only very basic or below basic quantitative skills (Kirsch, Jungeblut, Jenkins, & Kolstad, 2002). Respondents have difficulty with such tasks as identifying and integrating numbers in a lengthy text or performing two or more sequential steps to reach a solution. Although recent surveys have reported some improvement, a significant percentage of Americans continue to have below basic quantitative skills (22% in the 2003 National Assessment of Adult Literacy [NAAL], sponsored by the National Center for Education Statistics; Kutner, Greenberg, Jin, & Paulsen, 2006; for international comparisons, see Reyna & Brainerd, 2007).

Furthermore, it is not just the general population that has difficulty with numerical tasks. Studies have shown that even highly educated laypersons and health professionals have an inadequate understanding of probabilities, risks, and other chance-related concepts (Estrada, Barnes, Collins, & Byrd, 1999; Lipkus, Samsa, & Rimer, 2001; Nelson et al., 2008; Reyna, Lloyd, & Whalen, 2001; Sheridan & Pignone, 2002). These difficulties are reflected in poor risk estimation regardless of presentation format (i.e., in percentages or survival curves; Lipkus et al., 2001; Weinstein, 1999), improper calculation of the implications of diagnostic test results for disease probability (Reyna, 2004; Reyna & Adam, 2003), and inconsistent treatment decisions when outcomes are expressed in terms of absolute versus relative risk reduction (Forrow, Taylor, & Arnold, 1992). When surveyed, physicians generally indicate that it is important to provide quantitative risk estimates to their patients. However, they also report feeling more comfortable providing verbal estimates of risk than numerical ones, perhaps because of a lack of confidence and knowledge concerning the quantitative risk estimates or because they are aware that patients do not understand such estimates (Gramling, Irvin, Nash, Sciamanna, & Culpepper, 2004). Before we discuss the extent and ramifications of low numeracy, however, it is important to consider the fundamental question of how numeracy has been defined.

Defining Health Numeracy

Broadly defined, as we have noted, numeracy is the ability to understand and use numbers. Within this broad definition, however, numeracy is a complex concept, encompassing several functional elements. At the most rudimentary level, numeracy involves an understanding of the real number line, time, measurement, and estimation. Fundamental skills associated with numeracy include the ability to perform simple arithmetic operations and compare numerical magnitudes. At a higher level, numeracy encompasses basic logic and quantitative reasoning skills, knowing when and how to perform multistep operations, and an understanding of ratio concepts, notably fractions, proportions, percentages, and probabilities (Montori & Rothman, 2005; Reyna & Brainerd, 2008).

Educators and researchers have defined numeracy in various ways that reflect differences in their domains of study (see Table 1). The word *numeracy* was coined in 1959 by Geoffrey Crowther of the U.K. Committee on Education in the context of educating English schoolchildren. In its original sense, numeracy encompassed higher level mathematical reasoning skills that extended far beyond the ability to perform basic arithmetical operations (G. Lloyd, 1959):

> There is the need in the modern world to think quantitatively, to realize how far our problems are problems of degree even when they appear as problems of kind. Statistical ignorance and statistical fallacies are quite as widespread and quite as dangerous as the logical fallacies which come under the heading of illiteracy. (pp. 270–271)

Advancing a similarly expansive conception of numeracy, Paulos (1988) brought popular attention to the pervasive impairments in everyday functioning created by “innumeracy,” which he described as mathematical illiteracy. He emphasized the “inability to deal comfortably with the fundamental notions of number and chance” (p. 3), as well as difficulties in apprehending the magnitudes of extremely large and small numbers.

The concept of numeracy is often subsumed within the broader concept of literacy (Davis, Kennen, Gazmararian, & Williams, 2005). Experts have recognized that literacy is multifaceted and extends beyond simply reading and writing text to include mathematical reasoning and skills. Numeracy has thus been referred to as *quantitative literacy*, or “the ability to locate numbers within graphs, charts, prose texts, and documents; to integrate quantitative information from texts; and to perform appropriate arithmetical operations on text-based quantitative data” (Bernhardt, Brownfield, & Parker, 2005, p. 6). The conception of literacy as a multidimensional construct, and of numeracy as an integral subcomponent of literacy, is evinced by how the U.S. Department of Education defines literacy in its national literacy surveys, such as the National Adult Literacy Survey (NALS; Kirsch, Jungeblut, Jenkins, & Kolstad, 2002) and the NAAL (Kutner et al., 2006). In these surveys, literacy is a composite construct consisting of prose literacy (understanding and using information from texts), document literacy (locating and using information in documents), and quantitative literacy (applying arithmetical operations and using numerical information in printed materials).

Numeracy in the health context is often referred to as *health numeracy* and similarly conceptualized as a subcomponent of health literacy. As defined by Baker (2006), health literacy is an ordered skill set underlying the ability to understand written health information and to communicate orally about health. Baker’s definition includes prose, document, and quantitative literacy, as others do, but also “conceptual knowledge of health and health care” (p. 878). Quantitative literacy is assumed to be critical in these definitions because numbers—either in text or graphic format—pervade nearly all aspects of health communication. Other broad definitions of health literacy that have been proposed by various organizations include

quantitative reasoning skills as an integral component, in addition to basic computational skills and knowledge (see Table 1).

Health numeracy, however, is itself a broad concept because numerical reasoning in the health domain involves several different tasks and skills. One important task is to judge the relative risks and benefits of medical treatments; this task requires the ability to assess risk magnitude, compare risks, and understand decimals, fractions, percentages, probabilities, and frequencies, as these are the formats in which risk and benefit information is most often presented (Bogardus, Holmboe, & Jekel, 1999; Burkell, 2004). Other important tasks include interpreting and following medical treatment plans and navigating the health care system; such tasks require lower level, but still critical, numerical abilities including interpreting and following directions on a medication prescription label, scheduling follow-up medical appointments, and completing health insurance forms (Parker, Baker, Williams, & Nurss, 1995). Thus, health numeracy refers to various specific aspects of numeracy that are required to function in the health care environment (see Table 1). It is not simply the ability to understand numbers but rather to apply numbers and quantitative reasoning skills in order to access health care, engage in medical treatment, and make informed health decisions.

In an effort to develop an overarching framework for health numeracy that incorporates the varied skills that we have discussed, Golbeck, Ahlers-Schmidt, Paschal, and Dismuke (2005) conceptualized health numeracy as falling into four categories: basic (the ability to identify and understand numbers, as would be required to identify the time and date on a clinic appointment slip), computational (the ability to perform simple arithmetical calculations, such as calculating the number of calories from fat in a food label), analytical (the ability to apply higher level reasoning to numerical information, such as required to interpret graphs and charts), and statistical (the ability to apply higher level biostatistical and analytical skills, such as required to analyze the results of a randomized clinical trial). These four categories together compose the first level of Ancker and Kaufman's (2007) conceptual model.

As in Baker's (2006) approach, Ancker and Kaufman's (2007) model incorporates elements beyond the level of individuals' skills, most especially the health care environment. They proposed that health numeracy, or "the effective use of quantitative information to guide health behavior and make health decisions" (p. 713), depends on the interaction of three variables: (a) the individual-level quantitative, document, prose, and graphical literacy skills of the patient and provider; (b) the oral communication skills of both patient and provider; and (c) the quality and ease of use of information artifacts (such as decision aids and websites). Schapira et al. (2008) also described numeracy as a multifaceted construct that incorporates more than individuals' skills to include interpretive components influenced by patient affect.

The definitions that we have discussed introduce useful distinctions, such as contrasting basic computational versus reasoning abilities, and they are designed to highlight aspects of numeracy that have practical importance in the health care setting. However, none of the definitions is derived from an empirically supported theory of mathematical cognition. As we discuss, assessments of numeracy are similarly uninformed by theory. Assessments, in fact, are more narrowly construed than definitions of numeracy. Although conceptual definitions of health numeracy have stressed the health care environment, assessments have focused squarely on the skills of individuals, as we discuss in the following section.

Assessing Numeracy: National and International Surveys

How proficient are U.S. residents at understanding and working with numbers? Several national and international surveys of mathematical achievement suggest that although most Americans graduate from high school with basic mathematical skills, they are not proficient

and compare unfavorably with residents of other countries (Reyna & Brainerd, 2007). Moreover, most 12th graders lack skills that are essential for health-related tasks, falling short of what Golbeck et al. (2005) would describe as the analytical level at which numbers are used and understood. The National Assessment of Educational Progress (NAEP), or nation's report card, provides a comprehensive assessment of mathematical knowledge and skills. The NAEP comprises two types of assessments: a long-term trend assessment that has charted performance since 1973 and a "main" assessment that is periodically updated. In the most recent trends assessment, the average score for 12th-grade students was not appreciably different from the average score of 12th graders in 1973 (Perie, Moran, & Lutkus, 2005). Therefore, despite the increasing amount and complexity of health-related numerical information, students enter young adulthood no better prepared to process it than they were a generation ago.

The 2007 main NAEP assessed understanding of mathematical concepts and application of those concepts to everyday situations (Lee, Grigg, & Dion, 2007). Content areas included number properties and operations, measurement, geometry, data analysis and probability, and algebra. Achievement level was classified as basic (demonstrating partial mastery of grade-level skills), proficient (demonstrating solid grade-level performance), or advanced (demonstrating superior performance). The most recent data for 12th-grade mathematics performance were obtained from a nationally representative sample of more than 9,000 high school seniors (Grigg, Donahue, & Dion, 2007). Fully 41% of students performed at a below-basic level, 37% performed at a basic level, 20% performed at a proficient level, and 2% performed at an advanced level. This means that a substantial proportion of 12th graders did not have the basic mathematical skills required to, for example, convert a decimal to a fraction. In a theme that echoes across multiple national assessments, scores differed among subgroups. For example, Asian and Caucasian students performed better than African American, Hispanic, and American Indian students.

Similar findings were reported for the 2003 Program for International Student Assessment (PISA), which assesses mathematical literacy and problem-solving skills. Questions on the PISA reflect real-world situations requiring mathematical skills (e.g., converting currency for a trip abroad) and, thus, might be expected to be especially relevant to health numeracy. Like Golbeck et al.'s (2005) analytical level, PISA's emphasis is on using numerical knowledge and skills. In 2003 the performance of U.S. students was mediocre compared with that of students from other nations, with U.S. students scoring significantly below their peers in 23 countries. Average scores on each of the four mathematical literacy subscales (space and shape; change and relationships; quantity; and uncertainty) were significantly below the average scores for industrialized countries. Americans lagged behind their peers in mathematical problem solving as well: They ranked 29th of 39 countries tested and again scored significantly below the average for industrialized nations (although difficulties with mathematics spanned international borders; Lemke et al., 2004).

Not surprisingly, the mathematical proficiency of adults, as assessed by national surveys, is also lacking. The NALS, first carried out in 1992, surveyed a nationally representative sample of more than 26,000 adults (Kirsch et al., 2002). Each of the three literacy scales—prose, document, and quantitative—is divided into five proficiency levels. Twenty-two percent of adults performed at the lowest level of quantitative literacy, indicating that a substantial portion of the population has difficulty performing simple arithmetical operations. Twenty-five percent of adults performed at the next lowest level, which requires the ability to locate numbers and use them to perform a one-step operation. Nearly half the adult U.S. population could not identify and integrate numbers in a lengthy text or perform a numerical task requiring two or more sequential steps. Therefore, many adults lack the skills necessary to read a bus schedule to determine travel time to a clinic appointment or to calculate dosage of a child's medication based on body weight according to label instructions.

The 2003 NAAL (http://nces.ed.gov/NAAL/index.asp?file=KeyFindings/Demographics/Overall.asp&PageId=16), the most comprehensive assessment of the nation's literacy since the NALS, measured the literacy of a nationally representative sample of approximately 19,000 adults (Kutner et al., 2006). Included in this assessment was a scale designed to measure health literacy. Like the NALS, the NAAL evaluated prose, document, and quantitative literacy, and test items reflected tasks that people would likely encounter in everyday life. Adults were classified according to four literacy levels: below basic, basic, intermediate, and proficient. Those individuals functioning at the below-basic level would be expected to have only the simplest skills, such as being able to add two numbers, whereas those at a basic level would be expected to be able to perform simple, one-step arithmetical operations when the operation was stated or easily inferred. At a more advanced intermediate level, adults should be able to locate and use less familiar numerical information and solve problems when the operation is not stated or easily inferred. Overall, 36% of adults, or more than 93 million people, are estimated to perform at a below-basic or basic level. Those who scored lower on prose or document literacy also tended to score lower on quantitative literacy, but quantitative items elicited the lowest level of performance: Significantly more adults scored in the below-basic level on the quantitative scale (22%) than on the prose scale (14%) or document scale (12%; Kutner et al., 2006).

Subgroup analyses provide an even more disturbing picture of the nation's health literacy (Gonzales et al., 2004; Kutner et al., 2006; Lemke et al., 2004; Perie, Grigg, & Dion, 2005; Perie, Moran, & Lutkus, 2005; Reyna & Brainerd, 2007). Vulnerable subgroups with traditionally lower access to health care were, unfortunately, those with the lowest scores: Poverty and being a nonnative speaker of English were associated with lower scores. Among racial and ethnic subgroups, Hispanics and African Americans had the lowest average health literacy: Sixty-six percent of Hispanics and 58% of African Americans performed at a below-basic or basic level of health literacy. Adults age 65 and older had lower health literacy than younger adults: More than half the adults in the oldest age group had below-basic or basic health literacy. The latter figures are noteworthy in the context of health numeracy because older adults are more likely to have health problems. Although, as we have discussed, high school students performed poorly, adults who did not graduate from high school were worse off than those who did. Nearly one half of adults who did not complete high school functioned at a below-basic level.

In sum, representative national assessments of mathematical performance indicate that a slim majority of Americans have basic knowledge and skills. Performance of 12th graders has not changed in decades, despite rising requirements for numeracy. National performance levels for adults in mathematics generally raise questions that are borne out by low performance in assessments of health literacy, most notably quantitative literacy. This concern is heightened when we consider that millions of Americans score below average and that differences in performance are found across racial, ethnic, and socioeconomic groups. People with more health problems, and who had fewer resources to draw on to deal with those problems, had the lowest scores: Older, poorer, and less educated adults had lower health literacy than their younger, richer, and more educated counterparts. Thus, national assessments of mathematics achievement, of quantitative problem-solving performance, and of health literacy (including quantitative health literacy or numeracy) suggest that the average person is poorly equipped to process crucial health messages and medical information.

Assessing Health Numeracy: Specific Instruments

A variety of instruments have been developed that specifically assess health numeracy. These instruments are typically used in research studies and are not administered to nationally

representative samples. They do allow, however, a more fine-grained and formal assessment of mathematical skills. Without such assessment, it is difficult to determine whether an individual is sufficiently literate and numerate to function effectively in the health care environment (Nelson et al., 2008).

One reason is that physicians' ability to identify low-literate patients is limited. In three studies conducted at university-based medical clinics, physicians overestimated their patients' literacy skills (Bass, Wilson, Griffith, & Barnett, 2002; Lindau et al., 2002; Rogers, Wallace, & Weiss, 2006). Simply asking patients about their skills is unlikely to be useful because of the shame and stigma associated with low literacy and numeracy (Marcus, 2006; Parikh, Parker, Nurss, Baker, & Williams, 1996). Moreover, even if patients were willing, it is unlikely that self-assessments would be accurate (Dunning, Heath, & Suls, 2004). According to the NALS, most of the adults who performed at the lowest literacy level felt that they could read "well" and did not consider reading to be a problem (Kirsch et al., 2002). Similarly, Sheridan, Pignone, and Lewis (2003) found that although 70% of subjects perceived themselves to be good with numbers, only 2% answered three numeracy questions correctly.

As we discussed earlier in the context of national surveys, health numeracy often lags behind literacy. Thus, educational attainment does not ensure grade-level skills, and this is particularly true for mathematical skills (Doak & Doak, 1980; Kicklighter & Stein, 1993; McNeal, Salisbury, Baumgardner, & Wheeler, 1984; Rothman et al., 2006; Safeer & Keenan, 2005; Sentell & Halpin, 2006). Educated, literate people have difficulty understanding important numerical concepts such as relative risk reduction, number needed to treat, and conditional probabilities (e.g., probability of disease given a genetic mutation; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Reyna et al., 2001; Weiss, 2003). A surprising number of such people also have difficulty with elementary numerical concepts, such as whether a .001 risk is larger or smaller than 1 in 100 (Reyna & Brainerd, 2007). Therefore, although educational attainment is correlated with prose, document, and quantitative literacy, years of schooling cannot be assumed to translate into levels of numeracy (Rothman, Montori, Cherrington, & Pigone, 2008). The findings we have discussed—that providers cannot reliably identify patients with low numeracy, self-report is suspect, and level of education is misleading—indicate that specific instruments that assess numeracy are required.

Given the need for assessment of numeracy, it is not clear what form such assessment should take. Extant health numeracy measures can be broadly classified as either objective (respondents make numerical judgments or perform calculations, and their performance is evaluated objectively) or subjective (respondents express their level of confidence in their numerical ability). Objective measures ascertain a variety of abilities, such as how well people perform arithmetical operations, convert from one metric to another (e.g., express a frequency as a percentage), understand probability, and draw inferences from quantitative data. Subjective measures, which were conceived of as a less stressful and intimidating way to estimate level of numeracy, assess people's perceptions of their numerical competence (Fagerlin, Zikmund-Fisher, et al., 2007). Objective numeracy measures can be further subdivided into those that assess numeracy only or those that assess both literacy and numeracy, and into general or disease-specific measures. Numeracy measures have been related to measures of cognition, behaviors, and outcomes. In this section, we describe test characteristics and discuss relations to other measures in a subsequent section. We begin with objective composite measures that incorporate separate dimensions of competence, that is, literacy and numeracy.

Measures That Assess Multiple Dimensions

The Test of Functional Health Literacy in Adults (TOFHLA), the only health literacy measure to explicitly incorporate a numeracy component, represents a disease-general and composite

measure in that it tests both reading comprehension and numeracy separately (Davis et al., 2005; see Table 2). The TOFHLA reading comprehension section tests how well people understand instructions for a surgical procedure, a Medicaid application form, and an informed-consent document. The TOFHLA numeracy items pertain to tasks commonly encountered in health settings. They test the ability to follow instructions on a prescription medicine label, judge whether a blood glucose value is within normal limits, interpret a clinic appointment slip, and determine eligibility for financial assistance based on income and family size.

Although the TOFHLA tests reading comprehension and numeracy separately, it evaluates these sections psychometrically as a single unit. This feature, as well as the validation performed on the instrument, limits the utility of the TOFHLA for ascertaining numeracy per se. For example, concurrent validity of the TOFHLA was tested by correlating the TOFHLA with the Rapid Estimate of Adult Literacy in Medicine (REALM; Davis et al., 1991) and the reading subtest of the revised Wide Range Achievement Test (Jastak & Wilkinson, 1984), both of which test the ability to read and pronounce words (see Table 2). The numeracy section of the TOFHLA was not validated against a recognized measure of mathematical ability, such as the mathematics subtest of the Wide Range Achievement Test. Although reliability of a related measure in dentistry (TOFHLiD; Gong et al.) was determined for the reading comprehension and numeracy sections separately, construct validity was assessed with two reading tests, the REALM and the REALD-99, and the TOFHLA. The numeracy section of the TOFHLiD was also not validated against a specific numeracy measure.

Despite these limitations, the TOFHLA provides an indirect measure of key numeracy skills that contribute to functional health literacy (Parker et al., 1995). However, one drawback is that the TOFHLA can take up to 22 min to administer; for this reason, a short version (S-TOFHLA) containing two prose passages and four numeracy items, and requiring 12 min to administer, was developed (Baker, Williams, Parker, Gazmararian, & Nurss, 1999). Although the S-TOFHLA had adequate internal consistency (Cronbach's alphas for the numeracy and prose sections were .68 and .97, respectively) and was significantly correlated with the REALM (.80), the correlation of the numeracy items with the REALM (.61) was considerably lower than that of the reading comprehension section with the REALM (.81). Because the two prose passages of the S-TOFHLA were highly correlated with the full TOFHLA (.91), the four numeracy items were deleted. The S-TOFHLA was thus reduced to 36 reading comprehension items that required only 7 min to administer. The TOFHLA and original S-TOFHLA have been used to assess health literacy and numeracy in a range of health studies, including studies of geriatric retirees (Benson & Forman, 2002), Medicare patients (Baker, Gazmararian, Sudano, et al., 2002; Gazmararian et al., 1999; Gazmararian et al., 2003; Scott, Gazmararian, Williams, & Baker, 2002; Wolf et al., 2005), community-dwelling patients (Baker, Gazmararian, Sudano, & Patterson, 2000; Montalto & Spiegler, 2001), rheumatoid arthritis patients (Buchbinder, Hall, & Youd, 2006), spinal cord injury patients (Johnston, Diab, Kim, & Kirshblum, 2005), HIV-infected patients (Kalichman, Ramachandran, & Catz, 1999; Mayben et al., 2007), cardiovascular disease patients (Gazmararian et al., 2006), chronic disease patients (Williams, Baker, Parker, & Nurss, 1998), public hospital patients (Baker et al., 1997; Nurss et al., 1997; Parikh et al., 1996; Williams et al., 1995), emergency department patients (Baker et al., 1998), and Veterans Administration hospital patients (Artinian, Lange, Templin, Stallwood, & Hermann, 2003).

Like composite measures, integrative measures incorporate multiple dimensions of verbal and numerical processing (see Table 2). However, integrative measures involve tasks that require multiple skills for successful performance, such as both reading comprehension and numeracy. Unlike composite measures, literacy, numeracy, or other subscale scores cannot be separated; a single overall score is assigned. For example, in the Newest Vital Sign and the Nutrition Label Survey, people view a nutrition label and answer questions that require reading

comprehension skills as well as arithmetical computational and quantitative reasoning skills. They test both document literacy and quantitative literacy, respectively, in that they require the ability to search for and "use information from noncontinuous texts in various formats" and the ability to use "numbers embedded in printed materials" (Kutner et al., 2006, p. iv).

To complete the tasks on the Newest Vital Sign and the Nutrition Label Survey measures, people must be able to read and identify numbers contained in nutrition labels, ascertain which numbers are relevant to the specific question, determine the arithmetical operation required, and apply that operation. For example, in the Nutrition Label Survey, people are asked to view a soda nutrition label and determine how many grams of total carbohydrate are contained in a bottle. To answer this question, people must find where total carbohydrate content is listed on the label, determine the total carbohydrate content per serving (27 g), determine the number of servings per container (2.5), and apply the appropriate arithmetical operation to yield the correct answer (67.5 g; Rothman et al., 2006). Although such integrative tests involve realistic tasks, it is impossible to determine how much numeracy contributes to overall performance, and their reliability is lower than other measures (see Table 2).

Disease-General Numeracy Measures

A major shortcoming of existing composite and integrative scales is that they do not assess understanding of risk and probability. Adequate understanding of risk and probability is critical for decision making in all domains of health care, ranging from disease prevention and screening to treatment to end-of-life care (Nelson et al., 2008; Reyna & Hamilton, 2001). Risks and probabilities are examples of ratio concepts in mathematics (Reyna & Brainerd, 1994). After surveying performance on national and international assessments, Reyna and Brainerd (2007) concluded that ratio concepts such as fractions, percentages, decimals, and proportions are especially difficult to understand, and most adults perform poorly on these items. The National Mathematics Advisory Panel (2008) recently reached a similar conclusion after reviewing over 16,000 published studies of mathematics achievement. In addition to the processing complexities inherent in ratio concepts, risks and probabilities are associated with challenging abstract concepts such as chance and uncertainty (which refers to ambiguity as well as probability; Politi, Han, & Col, 2007).

One of the first efforts to assess people's understanding of risk information was undertaken by Black, Nease, and Tosteson (1995), who assessed numeracy by asking participants how many times a fair coin would come up heads in 1,000 tosses. Respondents were considered numerate if they answered the question correctly and provided logically consistent responses to other questions regarding the probability of developing or dying from breast cancer (e.g., estimating the probability of acquiring a disease as being greater than or equal to the probability of dying from the disease). Many numeracy measures feature such class-inclusion judgments (i.e., some probabilities are nested within other, more inclusive probabilities), a fact that has theoretical significance and, consequently, is discussed below in Theories of Mathematical Cognition: Psychological Mechanisms of Numeracy.

Another simple numeracy measure was developed by Weinfurt and colleagues (Weinfurt et al., 2003, 2005), who used a single question to assess how well patients understood the relative frequency of benefit from a treatment. They asked 318 oncology patients the meaning of the statement "This new treatment controls cancer in 40% of cases like yours" in the context of a physician's prognosis. Seventy-two percent of patients indicated that they understood this meant that "for every 100 patients like me, the treatment will work for 40 patients." However, 16% of patients interpreted this statement to mean either that the doctor was 40% confident that the treatment would work or that the treatment would reduce disease by 40%, and 12% of patients indicated that they did not understand the statement (Weinfurt et al., 2005).

Similarly, the L. M. Schwartz, Woloshin, Black, and Welch (1997) three-item numeracy scale tests familiarity with basic probability and related ratio concepts (e.g., proportions), which represents a departure from the TOFHLA's emphasis on simple arithmetical operations, basic understanding of time, and the ability to recognize and apply numbers embedded in text. The Schwartz et al. measure tests understanding of chance ("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?") and the ability to convert a percentage to a frequency (e.g., "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?"), and vice versa.

In their original study, L. M. Schwartz et al. (1997) examined the relationship of general numeracy to the ability to understand the benefits of screening mammography among 287 female veterans. Although 96% of the participants were high school graduates, more than half the women answered no or one question correctly, and only 16% answered all three questions correctly. Compared with less numerate women, more numerate women were better able to understand risk reduction information. This numeracy assessment was subsequently used to examine the relationship of numeracy to the validity of utility assessment techniques (Woloshin, Schwartz, Moncur, Gabriel, & Tosteson, 2001). Compared with low-numerate women, high-numerate women provided more logically consistent utility scores.

Similar findings were reported by S. R. Schwartz, McDowell, and Yueh (2004), who used a slightly modified version of the numeracy assessment to examine the effect of numeracy on the ability of head and neck cancer patients to provide meaningful quality of life data as measured by utilities for different states of health. Compared with low-numerate patients, high-numerate patients demonstrated greater score consistency on utility measures. Sheridan and Pignone (2002) also administered the L. M. Schwartz et al. (1997) numeracy assessment to 62 medical students and found that numeracy was associated with the ability to accurately interpret quantitative treatment data. The Schwartz et al. numeracy assessment has been used in an adapted or expanded format in other health research contexts (Estrada et al., 1999; Estrada, Martin-Hryniewicz, Peek, Collins, & Byrd, 2004; Parrott, Silk, Dorgan, Condit, & Harris, 2005).

Lipkus et al. (2001) sought to extend the L. M. Schwartz et al. (1997) numeracy assessment and test it in a highly educated population. To expand the numeracy assessment, they added eight questions framed in a nonspecific health context to the original three-item measure. They made a minor change to one of the three Schwartz et al. scale items: Rather than assess understanding of probability in the context of flipping a fair coin, the question was phrased in terms of rolling a fair six-sided die. Each of the eight new questions referred generally to either a "disease" or an "infection." As with the Schwartz et al. measure, the new items required an understanding of probability and ratio concepts (i.e., working with fractions, decimals, proportions, percentages, and probability). For example, the following question taps understanding of percentages: "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?"

Like other general numeracy measures that we have reviewed, which share similar items, the Lipkus et al. (2001) numeracy scale has acceptable reliability, but extensive psychometric validation and national norming data are lacking. However, the reported correlations between this measure and health-relevant judgments, such as risk perceptions, support its validity (see Effect of Numeracy on Cognition, Behaviors, and Outcomes, below). In any case, this numeracy scale is instructive in that it clearly demonstrated that even college-educated people have difficulty with basic ratio concepts (i.e., probability, percentages, and proportions) and perform poorly when asked to make relatively simple quantitative judgments.

In fact, when one compares the performance of the less well educated participants in the L. M. Schwartz et al. (1997) study with that of the more highly educated participants in the Lipkus et al. (2001) study, the results are remarkably similar. In the Schwartz et al. study, 36% of the 287 participants had some college education, compared with 84%–94% of the 463 participants in the Lipkus et al. study. Despite this difference in educational attainment, 58% of the participants in both studies answered no or one question correctly. Sixteen percent of subjects in the Schwartz et al. study and 18% in the Lipkus et al. study answered all the questions correctly. (As these comparisons suggest, studies that control for effects of education, income, and other factors have shown that numeracy accounts for unique variance—e.g., Apter et al., 2006; Cavanaugh et al., 2008—though controls for ethnicity and socioeconomic status are inconsistent.) It is troubling that even college-educated people have difficulty with ratio concepts because ratio concepts are critical for understanding and interpreting risk, which in turn is required to make effective medical judgments (Reyna, 2004).

Building on the Lipkus et al. (2001) numeracy scale, Peters and colleagues (Greene, Peters, Mertz, & Hibbard, 2008; Hibbard, Peters, Dixon, & Tusler, 2007; Peters, Dieckmann, Dixon, Hibbard, & Mertz, 2007) added four items to create an expanded numeracy scale. The new items, which make the Lipkus et al. numeracy scale more challenging, test familiarity with ratio concepts and the ability to keep track of class-inclusion relations (Barbey & Sloman, 2007; Reyna, 1991; Reyna & Mills, 2007a). For example, the following question from this test requires processing of nested classes and base rates and then determining the positive predictive value of a test (i.e., the probability that a positive result indicates disease):

> Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. The table below summarizes all this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor? (Peters, Dieckmann, et al., 2007, p. 174)

The correct answer is .50.

The Medical Data Interpretation Test calls on even more advanced skills, compared with the general numeracy scales just reviewed (L. M. Schwartz, Woloshin, & Welch, 2005; Woloshin, Schwartz, & Welch, 2007). Whereas most general numeracy measures assess a range of arithmetic computation skills, basic understanding of probability and risk, and simple quantitative reasoning skills, the Medical Data Interpretation Test "examines the ability to compare risks and put risk estimates in 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)" (L. M. Schwartz et al., 2005, p. 291). The instrument tests skills needed to interpret everyday health information, such as information contained in drug advertisements or health-related news reports. In addition to the skills needed to complete the other general numeracy measures, the Medical Data Interpretation Test requires a more sophisticated understanding of base rates, absolute risk, relative risk, knowledge of the kinds of information needed to assess and compare risks, and the ability to apply inferential reasoning to health information. The test also taps understanding of epidemiological concepts and principles, such as incidence, the distinction between population-level and individual-level risk, and clinical trial design (e.g., why comparison groups are needed for clinical trials).

For example, one of the test questions pertains to a description of a clinical trial of a new drug for prostate cancer. In this trial, only three subjects taking the study drug developed prostate

cancer. On the basis of this information, the test taker is asked to select the most critical question for understanding the results of the clinical trial from among these options: (a) Who paid for the study? (b) Has the drug been shown to work in animals? (c) What was the average age of the men in the study? (d) How many men taking the sugar pill developed prostate cancer? Another series of questions tests reasoning skills. People are first asked to estimate a person's chance of dying from a heart attack in the next 10 years and then to estimate that same person's chance of dying for any reason in the next 10 years. To answer correctly, a person would need to recognize that the risk of dying from all causes is greater than the risk of dying from a single cause (another class-inclusion judgment; Reyna, 1991). The Medical Data Interpretation Test has been translated into Dutch and validated among Dutch university students (Smerecnik & Mesters, 2007). Like other disease-general numeracy measures, the Medical Data Interpretation Test has face validity, as it seems to require skills involved in medical decisions.

In sum, performance on several disease-general numeracy tests has been linked to health-related risk perceptions, understanding of treatment options, measurement of patient utilities and other relevant cognitions, behaviors, and outcomes. However, as we discuss in greater detail below, the content of the tests has been determined by using prior measures and commonsense assumptions; none of these tests explicitly taps research or theory in mathematical cognition.

Disease-Specific Numeracy Measures

Disease-specific numeracy instruments have also been developed to assist with the management of chronic conditions that require self-monitoring (see Table 2). These tests have yet to garner the extent of empirical support that disease-general measures have, but they allow researchers (and potentially clinicians) to focus on skills relevant to specific diseases and treatment regimens. Apter et al. (2006) developed a four-item numeracy questionnaire that assesses understanding of basic numerical concepts required for asthma self-management. The questionnaire tests a patient's understanding of basic arithmetic (e.g., determining how many 5-mg tablets are needed if your daily dose of prednisone is 30 mg) and percentages, as well as the ability to calculate and interpret peak flow meter values. Estrada et al. (2004) expanded the L. M. Schwartz et al. (1997) three-item numeracy assessment to test the ability of patients taking warfarin (an anticoagulant) to handle basic numerical concepts needed for anticoagulation management. They added three items that assess basic knowledge of addition, subtraction, multiplication, and division that apply specifically to warfarin (e.g., "You have 5 mg pills of Coumadin [warfarin] and you take 7.5 mg a day. If you have 9 pills left, would you have enough for one week?"). Finally, the Diabetes Numeracy Test (Cavanaugh et al., 2008;Huizinga et al., 2008) is a 43-item instrument that taps multiple numeracy domains relevant to diabetes nutrition, exercise, blood glucose monitoring, oral medication use, and insulin use. An abbreviated 15-item version of the Diabetes Numeracy Test, which demonstrates a .97 correlation with the 43-item instrument, is also available (Huizinga et al., 2008; see also Montori et al., 2004). Although of recent vintage, these disease-specific numeracy scales show promise in predicting medical outcomes that are tied to measured skills, as discussed further below (Estrada et al., 2004).

Subjective Numeracy Measures

Unlike the objective measures of numeracy that we have reviewed so far, subjective numeracy measures attempt to assess how confident and comfortable people feel about their ability to understand and apply numbers without actually having to perform any numerical operations. A primary rationale underlying researchers' interest in subjective measures has been to increase the feasibility and acceptability of measuring numeracy for respondents, because objective measures are arduous and potentially aversive. The aim has been to develop a measure that would allow subjective numeracy to be used as a proxy for objective numeracy. The first

subjective numeracy measures to be developed were the STAT–Interest and STAT–Confidence scales, created by Woloshin et al. (2005) to assess people's attitudes toward health-related statistics. The three items on the STAT–Confidence scale cover perceived ability to understand and interpret medical statistics; the five items on the STAT–Interest scale pertain to level of attention paid to medical statistics in the media and in the medical encounter.

Although study participants reported generally high levels of interest and confidence in medical statistics, the interest and confidence scales were weakly correlated with a validated measure of objective numeracy, the Medical Data Interpretation Test (r = .26 and r = .15, respectively), suggesting that people are poor judges of their ability to use medical statistics. This finding is not entirely unexpected, as it is well documented that people tend to be poor judges of their abilities, particularly in the educational domain (Dunning et al., 2004). The ability to self-assess is subject to such systematic biases as unrealistic optimism, overconfidence, and the belief that one possesses above-average abilities. However, in contrast to the findings of Woloshin et al. (2005), the Subjective Numeracy Scale (Fagerlin, Zikmund-Fisher, et al., 2007; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007) demonstrated a moderate correlation (rs = .63–.68) with the Lipkus et al. (2001) numeracy scale, suggesting that subjective measures may be a potentially viable means of estimating numeracy. Naturally, the most persuasive evidence of validity for subjective measures would be that they are able to predict objective performance, but little evidence been gathered on this point (but see Fagerlin, Zikmund-Fisher, et al., 2007). Further research is also needed to determine the potential clinical utility of subjective measures such as these (Nelson et al., 2008).

Summary of Measures

In sum, to date various measures have been developed to assess health numeracy, yet no single measure appears to capture the totality of this construct. Rather, the objective health numeracy measures can be thought of as representing a continuum of competencies, ranging from rudimentary numeracy skills (such as the ability to tell time and perform one- and two-step arithmetic problems) to intermediate level skills (including the ability to apply basic ratio concepts involved in understanding risks and probabilities) to advanced numeracy skills requiring higher level inferential reasoning skills (such as the ability to determine the positive predictive value of a test). Examples of measures that test basic, low-level skills include the TOFHLA and TOFHLiD. Measures that fall between basic and intermediate (analytical) level skills include the Newest Vital Sign and the Nutrition Label Survey. All the general and disease-specific measures that we have examined require at least some intermediate-level skills. The Medical Data Interpretation Test and the Peters, Dieckmann, et al. (2007) expanded numeracy scale both require higher level reasoning skills to assess risk. Yet, as we discuss in the next section, progress in assessment has outpaced progress in basic understanding of numeracy on a causal level, that is, in understanding the cognitive mechanisms that underlie numeracy and how numeracy affects health behaviors and outcomes.

Effect of Numeracy on Cognition, Behaviors, and Outcomes

For clinicians and policymakers, the importance of numeracy in health care is not as an end in itself but as a means of achieving health behaviors and outcomes that matter for patients. Because effective health care depends so critically on adequate patient understanding, numeracy has the potential to affect a variety of important outcomes, ranging from health decision making, health services utilization, and adherence to therapy to more distal outcomes including morbidity, health-related quality of life, and mortality. As our subsequent review of this research details, there is evidence for the expected associations between numeracy and various cognitive milestones along the causal path to such outcomes, ranging from effects on comprehension to effects on judgment and decision making; in a few studies, associations with health behaviors and outcomes have been demonstrated. Figure 1 portrays some of the points

on this path. We begin with perceptions of risks and benefits, followed by measurement of patient utilities (e.g., values for health states such as disability as opposed to death), information presentation and formatting, and, last, health behaviors and medical outcomes.

Perceptions of Risk and Benefit

An understanding of the risks and benefits associated with particular choice options is important for many health decisions. For example, patients are expected to understand and weigh the risks and benefits of various treatment options shortly after being diagnosed with an illness. In this section, we review literature showing that people lower in numerical ability have consistent biases in their perceptions of risk and benefit.

Many of the studies examining risk perceptions have been conducted in the context of breast cancer research. Black et al. (1995) asked women between the ages of 40 and 50 (N = 145) several questions about the probability that they would develop or die of breast cancer in the next 10 years. They measured numeracy with a single question (the number of times a fair coin would come up heads in 1,000 tosses). The entire sample overestimated their personal risk of breast cancer, compared with epidemiological data, and those lower in numeracy made even larger overestimations than those higher in numeracy.

L. M. Schwartz et al. (1997) also asked women (N = 287) to estimate the risk of dying from breast cancer both with and without mammography screening. They presented the women with risk reduction information (i.e., risk reduction attributable to mammography) in four formats and calculated accuracy by how well they adjusted their risk estimates in light of the new information. After controlling for age, income, level of education, and the format of the information, they found that participants higher in numeracy were better able to use the risk reduction data to adjust their risk estimates.

In another study, Woloshin, Schwartz, Black, and Welch (1999) asked women (N = 201) to estimate their 10-year risk of dying from breast cancer as a frequency out of 1,000. In addition, they asked the women to estimate how their risk compared with that of an average woman their age. Numeracy was measured with the three-item scale used by Schwartz et al. (1997). After controlling for education and income, they found that numeracy was not related to participants' comparison judgments, but participants lower in numeracy overestimated their risk of dying from breast cancer in the next 10 years. This study showed that participants lower in numeracy might still be able to make accurate risk comparisons, even though they are not able to make unbiased risk estimates.

In another study of breast cancer risk (Davids, Schapira, McAuliffe, & Nattinger, 2004), a sample of women estimated their 5-year and lifetime risk of breast cancer and completed the L. M. Schwartz et al. (1997) scale. Similar to the findings above, participants (N = 254) as a whole overestimated their risk of breast cancer (compared with epidemiological data), with those lower in numeracy making larger errors in their estimates than those higher in numeracy (when controlling for age, race, education, and income). In a separate report, these authors also showed that numeracy was related to consistent use of frequency and percentage risk rating scales (Schapira, Davids, McAuliffe, & Nattinger, 2004). After controlling for age, health literacy, race, and income, they found that higher numeracy was shown to be predictive of using the percentage and frequency scales in a consistent manner (i.e., giving the same responses on both frequency and percentage scales for the 5-year and lifetime breast cancer risk estimates, respectively).

There have also been a few studies that have not found a relationship between numeracy and breast cancer risk estimates. Dillard, McCaul, Kelso, and Klein (2006) investigated whether poor numeracy skills could account for the finding that women consistently overestimate their

risk of breast cancer even after receiving epidemiological information about the risk. In this study ($N = 62$), numeracy as measured by the L. M. Schwartz et al. (1997) scale was not related to persistent overestimation of breast cancer risk. Another group of researchers asked a sample of Black and White women ($N = 207$) to estimate their 5-year survival after a diagnosis of breast cancer, along with an estimate of the relative risk reduction due to screening mammography (Haggstrom & Schapira, 2006). Also using the Schwartz et al. measure, they found no effect of numeracy after controlling for other demographic variables (e.g., race, family history of breast cancer, income, insurance type, and level of education). These null results amount to failures to detect relationships rather than evidence of their absence.

There have also been studies of perceptions of risks and benefits outside the breast cancer domain. In a large survey of cancer patients ($N = 328$), Weinfurt et al. (2003) asked participants to estimate the chances that they would benefit from an experimental cancer treatment. Numeracy was measured with a single multiple-choice question about a treatment that controlled cancer in "40% of cases like yours" (the correct answer: the treatment will work for 40 out of 100 patients like me). Patients who did not answer the numeracy question correctly perceived greater benefit from experimental treatment. In another study, participants were presented with several hypothetical scenarios that described a physician's estimate of the risk that a patient had cancer (Gurmankin, Baron, & Armstrong, 2004b). The authors asked participants to imagine that they were the patient described and to rate their risk of cancer. Numeracy was measured with a scale adapted from Lipkus et al. (2001). They found that patients lower in numeracy were more likely to overestimate their risk of cancer.

Similar results have been obtained in nonhealth domains. For example, Berger (2002) presented news stories describing an increase in burglaries, along with frequency information. Participants lower in numeracy were more apprehensive about the increase in burglaries. In addition, Dieckmann, Slovic, and Peters (2009) presented a narrative summary, along with numerical probability assessments, of a potential terrorist attack. Participants lower in numeracy reported higher perceptions of risk and recommended higher security staffing. Consonant with studies reviewed earlier, those lower in numeracy were less sensitive to numerical differences in probability and focused more on narrative evidence.

In conclusion, several studies have found that numeracy is related to perceptions of health-related risks and benefits. Participants lower in numeracy tend to overestimate the risk of cancer and other risks, are less able to use risk reduction information (e.g., about screening) to adjust their risk estimates, and may overestimate benefits of uncertain treatments. Note that low numeracy does not lead to randomly wrong perceptions of risks and benefits, as hypotheses about lack of skills or about imprecision might expect, but rather to systematic overestimation. Woloshin et al. (1999), among others, suggested that the form of the risk question may be part of the problem. In other words, participants low in numerical ability might have trouble expressing their risk estimates on the scales generally used in this domain (but see Reyna & Brainerd, 2008). As discussed, the low numerate seem to have difficulty using frequency and percentage risk scales consistently. However, difficulty using risk scales does not in itself predict overestimation. Instead, uncertainty about the meaning of numerical information, resulting from lower numeracy, may promote affective interpretations of information about risks (i.e., fearful interpretations) and about benefits (i.e., hopeful interpretations). Alternatively, overestimation may reflect the domains studied; cardiovascular risk, for example, might be underestimated for women because it is perceived to be a disease of men. Future research should focus on disentangling response scale effects from affective, motivational, and conceptual factors that may influence how those low in numeracy interpret risk and benefit information.

Eliciting Judgments of Value (Utilities)

Much research has focused on measuring the values, or utilities, that patients place on different health states and health outcomes. Obtaining reliable and valid assessments of how patients value different health states is important for modeling their decision making as well as improving health care delivery. The two primary methods for eliciting these utilities are the standard gamble and time-trade-off methods (Lenert, Sherbourne, & Reyna, 2001; Woloshin et al., 2001). Both methods require a participant to make repeated choices between different hypothetical health states until they are indifferent between options—namely, a choice in which they do not favor one health state over the other. For example, imagine that a patient with health problems has a life expectancy of 20 years. The patient is faced with a series of choices between living in his or her current state of health for the remainder of life or living with perfect health for some shorter amount of time. A utility for the patient's current health state can be calculated based on the point at which the patient finds both choices equally attractive (e.g., 10 years of perfect health vs. remainder of life with current health = 10/20; thus, utility = .5). The methods used to elicit utilities from patients often require them to deal with probabilities and/or make trade-offs between states. Because of the quantitative nature of the task, some researchers have questioned the validity of the standard approach to eliciting utilities and their dependence on the numerical abilities of patients.

Woloshin et al. (2001) used three methods for eliciting utilities about the current health of a sample of women ($N = 96$). The participants completed a standard gamble and a time-trade-off task and rated their current health on an analog scale, as well as completed the L. M. Schwartz et al. (1997) numeracy scale. Low-numerate participants showed a negative correlation between a question about current health and utilities generated from the standard gamble and time-trade-off tasks (i.e., valuing worse health higher than better health), indicating that these participants had difficulty with these quantitative utility elicitation tasks. The high-numerate participants showed the expected positive correlation (between self-reported health and utility for current health) for the same two tasks. It is interesting to note that all participants showed a positive correlation with the analog rating scale, which demanded less quantitative precision. Similar studies have been conducted with head and neck cancer patients. Utility scores were more consistent for the numerate patients, and their scores were more strongly correlated with observed functioning (S. R. Schwartz et al., 2004). These findings also suggest that the standard methods for assessing utility may be untrustworthy in patients with limited numerical ability. Similar conclusions have been reached when using a subjective numeracy measure (Zikmund-Fisher et al., 2007).

These studies indicate that patients lower in numeracy have difficulty with the standard procedures for assessing utilities, especially those that are more quantitatively demanding. Low-numerate patients have difficulties dealing with probabilities, but they also appear to have trouble making trade-offs. Making trade-offs between hypothetical states involves additional reasoning skills that do not seem to be necessary when simply comparing probabilities. Future work should investigate the interplay between the different skills that are needed to complete these tasks, which could lead to new methods of eliciting utilities that are appropriate for patients at all levels of numerical ability.

Information Presentation and Formatting

Given the gap between the intended meaning of health information and what people construe that information to mean, researchers have tried to identify optimal methods of presenting numerical information to improve understanding (e.g., Fagerlin, Ubel, Smith, & Zikmund-Fisher, 2007; Lipkus & Hollands, 1999; Maibach, 1999; Peters, Hibbard, Slovic, & Dieckmann, 2007; Reyna, 2008; Reyna & Brainerd, 2008). Several experiments have examined

how individuals varying in numerical ability are affected by information framing and by presenting probabilities in different formats.

Framing effects have proven to be a relatively robust phenomenon in psychological research (e.g., Kuhberger, 1998). For example, presenting the risk of a surgical operation as an 80% chance of survival versus a 20% chance of death (i.e., gain vs. loss framing) has been shown to change perceptions of surgery (McNeil, Pauker, Sox, & Tversky, 1982). In one study conducted by Stanovich and West (1998), participants with lower total Scholastic Aptitude Test (SAT) scores were found to be more likely to show a framing effect for risky choices about alternative health programs. (Because gain and loss versions of framing problems are mathematically equivalent, ideally, choices should be the same; framing effects occur when choices differ across frames.) In a more recent study, Stanovich and West (2008) found that the effect of frame type failed to interact with cognitive ability. If anything, inspection of the means revealed a trend in the wrong direction (the high-SAT group displayed a slightly larger framing effect); according to standard theories, higher ability participants should treat equivalent problems similarly and not show framing effects. Regarding these conflicting findings, Stanovich and West (2008) pointed out that within-subjects framing effects seem to be associated with cognitive ability but between-subjects effects are not. However, in these studies, effects were probably due to general intelligence rather than to numeracy because results for verbal and quantitative measures of cognitive ability (i.e., verbal and quantitative SAT scores) did not differ.

Controlling for general intelligence (using self-reported total SAT scores), Peters et al. (2006) compared framing effects for low- and high-numerate subjects ($N = 100$). These groups were defined based on a median split using their scores on the Lipkus et al. (2001) numeracy scale (low numeracy was defined as two to eight items correct on the 11-item scale). Naturally, college-student participants who were relatively less numerate were not necessarily “low” in numeracy in an absolute sense. Nevertheless, less and more numerate participants rated the quality of hypothetical students’ work differently when exam scores were framed negatively (e.g., 20% incorrect) versus positively (e.g., 80% correct). That is, less numerate participants showed larger framing differences. Peters and Levin (2008) showed in a later study that the choices of the more numerate were accounted for by their ratings of the attractiveness of each option in a framing problem (i.e., the sure thing and the risky option). The choices of the less numerate showed an effect of frame beyond the rated attractiveness of the options, demonstrating effects for both single-attribute framing (e.g., exam scores) and risky-choice framing.

Peters et al. (2006) also examined whether numerical ability affected the perception of probability information. Participants ($N = 46$) rated the risk associated with releasing a hypothetical psychiatric patient. Half read the scenario in a frequency format (“Of every 100 patients similar to Mr. Jones, 10 are estimated to commit an act of violence to others during the first several months after discharge”), and the other half received the same information in a percentage format (“Of every 100 patients similar to Mr. Jones, 10% are estimated to commit an act of violence to others during the first several months after discharge”). More numerate participants did not differ in their risk ratings between the two formats. Less numerate participants, however, rated Mr. Jones as being less of a risk when they were presented with the percentage format.

Peters, Dieckmann, et al. (2007) also explored the relationship between numeracy and the format of numerical information. In each experiment, participants ($N = 303$) were presented with measures of hospital quality and asked to make an informed hospital choice. Numeracy was measured with the Peters, Dieckmann et al. expanded numeracy scale. In the first study, participants saw a number of hospital-quality indicators (e.g., percentage of time that guidelines

for heart attack care were followed) as well as nonquality information (e.g., number of visiting hours per day) for three hospitals. Information about cost was also provided. Participants were asked to choose a hospital that they would like to go to if they needed care, and they also responded to a series of questions about the information presented (e.g., which hospital is most expensive?). The information was displayed in one of three ways: in an unordered fashion, with the most relevant information (cost and quality) presented first, or with only the cost and quality information displayed (other nonquality information was deleted). Those lower in numeracy showed better comprehension when information was ordered and when information was deleted, compared with the unordered condition. (Those higher in numeracy also benefited from deleting information.)

In a second study, participants were told to imagine that they needed treatment for heart failure and were asked to choose among 15 hospitals based on three pieces of information: cost, patient satisfaction, and death rate for heart failure patients. The formatting of information was varied by including black and white symbols or colored traffic light symbols to help participants evaluate the goodness or badness of each piece of information. The traffic light symbols were thought to be easier to understand. However, the low-numerate participants made better choices (i.e., used cost and quality indicators) with the "harder" black-and-white symbols compared with the "easier" traffic light symbols, whereas the reverse was true for the high-numerate participants—a result that is difficult to interpret. In a third hospital choice study, the authors found that low-numerate participants were particularly sensitive to the verbal framing of the information. Low-numerate participants showed greater comprehension when information was presented such that a higher number means better (the number of registered nurses per patient) compared with when a lower number means better (the number of patients per registered nurse).

Numeracy has also been related to reading graphs. Zikmund-Fisher et al. (2007) presented participants ($N = 155$) with a survival graph that depicted the number of people given two drugs who would be alive over a 50-year period. They then asked four questions about information displayed in the graph (e.g., regarding what year the difference in total survival between Pill A and Pill B was largest). They measured numeracy with a subjective numeracy measure. The ability to correctly interpret the survival graphs was strongly related to numeracy, with those higher in numeracy better able to interpret the graphs.

In another study, effects of format on trust and confidence in numerical information were examined. Gurmankin, Baron, and Armstrong (2004a) conducted a web-based survey in which they presented subjects ($N = 115$) with several hypothetical risk scenarios. The scenarios depicted a physician presenting an estimate of the risk that a patient had cancer in three formats (verbal, numerical probability as a percentage, or numerical probability as a fraction). Participants then rated their trust and comfort with the information, as well as whether they thought the physician distorted the level of risk. Numeracy was measured by adapting the Lipkus et al. (2001) scale. Even after adjusting for gender, age, and education, Gurmankin et al. found that those subjects with the lowest numeracy scores trusted the information in the verbal format more than the numerical, and those with the highest numeracy scores trusted the information in the numerical formats more than the verbal.

Sheridan and colleagues (Sheridan & Pignone, 2002; Sheridan et al., 2003) conducted two studies in which they assessed the relationship between numeracy and ability to interpret risk reduction information in different formats. Participants in both studies were presented with baseline risk information about a disease and then given risk reduction information in one of four formats: relative risk reduction, absolute risk reduction, number needed to treat, or a combination of all methods. Number needed to treat is an estimate of the number of patients who must be treated in order to expect that one patient will avoid an adverse event or outcome over a period. Mathematically, it is the inverse of absolute risk reduction (the decrease in

disease due to treatment) and was introduced because of difficulties in understanding other risk reduction formats. Numeracy was measured with items from the L. M. Schwartz et al. (1997) scale. In a sample of first-year medical students ($N = 62$) and a sample of patients from an internal medicine clinic ($N = 357$), participants lower in numeracy had more difficulty using the risk reduction information and, in particular, had trouble with the number-needed-to-treat format.

Finally, controlling for gender and ethnicity, Parrott et al. (2005) presented statistical evidence concerning the relationship between a particular gene and levels of LDL cholesterol. The statistical information was presented in either a verbal form with percentage information or a visual form that showed a bar graph of the mortality rates. They were interested in whether perceptions of the evidence differed between the formats and whether numeracy was related to these perceptions (four numeracy items were adapted from L. M. Schwartz et al., 1997). They did not find any relationships between numeracy and comprehension, perceptions of the quality of the evidence, or perceptions of the persuasiveness of the evidence. The authors noted, however, that the restricted range of numerical abilities may have contributed to the null effects.

In sum, low-numerate participants tend to be worse at reading survival graphs, more susceptible to framing effects in some experiments, more sensitive to the formatting of probability and risk reduction information, and more trusting of verbal than numerical information. Many of these studies do not control for general intelligence, although some do and still obtain effects of numeracy (e.g., Peters et al., 2006). Regardless of whether numeracy per se is the problem, those who score low on these assessments can be helped by presenting information in a logically ordered format and displaying only the important information, presumably decreasing cognitive burden. Additional research is needed to further elucidate the presentation formats that are most beneficial for individuals at different levels of numerical ability (but for initial hypotheses based on research, see Fagerlin, Ubel, et al., 2007; Nelson et al., 2008; Reyna & Brainerd, 2008). Variations in formatting have generally not been theoretically motivated; variations in numerical ability further complicate theoretical predictions about formatting. Effective prescriptions for formatting, therefore, await deeper understanding of the locus of effects of presentation format.

Health Behaviors and Outcomes

Given the research that we have reviewed summarizing the deleterious effects of low numeracy on perceptions of crucial health information and the vulnerability of low-numerate individuals to poor formatting of that information, it would not be surprising if numeracy were related to health behaviors and medical outcomes (see Figure 1). The limited data that are available support such a conclusion. As we have discussed in some detail, several studies have demonstrated a relationship between numeracy and disease risk perceptions, which are themselves known to be critical determinants of health behaviors (e.g., for reviews, see Brewer, Weinstein, Cuite, & Herrington, 2004;Klein & Stefanek, 2007;Mills, Reyna, & Estrada, 2008). Therefore, numeracy, through its effect on perceptions of risks and benefits, would be expected to change health behaviors and outcomes.

Consistent with this suggestion, studies that found low numeracy to be associated with a tendency to overestimate one's cancer risk also showed that such overestimation affected the perception of the benefits of cancer screening as well as screening behaviors, generally encouraging screening but, in the extreme, perhaps leading to fatalistic avoidance of screening (e.g., Black et al., 1995; Davids et al., 2004; Gurmankin et al., 2004b; L. M. Schwartz et al., 1997; Woloshin et al., 1999). These data support the conclusion that numeracy may influence distal health outcomes through effects on risk perceptions and other mediating processes, as shown in the model depicted in Figure 1. Notably, however, the data on the relationship between risk perceptions and outcomes have not always been consistent (see Mills et al.,

2008; Reyna & Farley, 2006). For example, numeracy was found to be unrelated to estimates of breast cancer survival and survival benefit from screening mammography in a study that controlled for the effect of sociodemographic variables (Haggstrom & Schapira, 2006).

As discussed above, studies have also explored the relationship between numeracy and people's ability to provide utility estimates of health states, another intermediate factor of importance in health-related decisions (see Figure 1). An important question is whether health utilities can serve as proxy measures for objective medical outcomes. Indeed, some who emphasize quality of life suggest that perceived utilities are superior to objective health states as ultimate measures of outcomes (a suggestion that we would not endorse). In any case, this conclusion does not follow if health utilities are inaccurate, as they are for those low in numeracy (e.g., S. R. Schwartz et al., 2004;Woloshin et al., 2001;Zikmund-Fisher et al., 2007). It is not known whether numeracy influences understanding of the information presented with these techniques, performance of the trade-off tasks themselves, or people's ability to communicate their preferences. However, the overall implication is that limited numeracy may interfere with patients' ability to express their preferences and clinicians' ability to elicit them. These factors also represent important potential mediators of the effects of low numeracy on intermediate health outcomes such as patient-centered communication and informed decision making.

Most of what we know and suspect about the health outcomes of numeracy is inferred from descriptive studies in the related and better developed domain of health literacy. A number of studies using the TOHFLA have linked health literacy to several important outcomes and provide indirect evidence of the effects of numeracy, as the TOHFLA measures both health literacy and functional quantitative skills. Health literacy as measured by the TOHFLA has been associated with poor knowledge and understanding of various chronic diseases—including hypertension, diabetes mellitus, congestive heart failure, and asthma—among patients with these conditions (Gazmararian et al., 2003; Williams, Baker, Honig, Lee, & Nowlan, 1998; Williams, Baker, Parker, & Nurss, 1998). Further down the potential causal path from patient understanding to health behaviors and outcomes (as depicted in Figure 1), health literacy as measured by the TOHFLA has also been associated with lower utilization of preventive medical services, such as routine immunizations, Pap smears, and mammograms (Scott et al., 2002), and lower adherence to highly active antiretroviral therapy in HIV patients (Kalichman et al., 1999).

Low numeracy per se has also been found to be associated with poor patient self-management of chronic disease. In a prospective study, Estrada et al. (2004) examined numeracy with respect to anticoagulation control among patients taking warfarin. This labor-intensive therapy requires patients to monitor quantitative laboratory test results and respond to these results by calculating and making adjustments in medication doses. As expected, low numeracy was found to be associated with poor anticoagulation control (ascertained in terms of the extent to which patients' laboratory test results were within the therapeutic target range). In a cross-sectional study, Cavanaugh et al. (2008) examined the association between numeracy and diabetes self-management skills using the Diabetes Numeracy Test (Cavanaugh et al., 2008). Higher diabetes-related numeracy was associated with higher perceived self-efficacy for managing diabetes. However, higher numeracy was only weakly associated with a key outcome, hemoglobin A_{1c}, a measure of glycemic control in diabetic patients.

If low numeracy leads to poor understanding of health information, which in turn leads to lower utilization of health services and poor treatment adherence and disease self-management, then the next expected outcome in the causal chain would be greater morbidity (see Figure 1). Controlling for covariates, descriptive studies using the TOHFLA add further indirect evidence of a numeracy effect. For example, Baker and colleagues (Baker, Gazmararian, Williams, et

al., 2002;Baker et al., 2004) found an association between low health literacy and greater utilization of emergency department services and risk of future hospital admission among urban patient populations. Health literacy has also been shown to be associated with lower self-rated physical and mental health functioning as measured by the Medical Outcomes Study 36-Item Short-Form Health Survey, or SF-36, which is predictive of both morbidity and mortality (Wolf et al., 2005). In one study of patients prescribed inhaled steroids for treatment of asthma, low numeracy was found to be correlated with a history of hospitalizations and emergency room visits for asthma (Apter et al., 2006). Thus, in combining results for disease management with direct and indirect evidence for disease outcomes (e.g., hospitalizations), there is limited evidence that low numeracy affects morbidity and mortality.

Unfortunately, studies have not examined outcomes such as long-term morbidity and mortality, clear needs for future research. Our review of assessments indicates that measures of literacy and numeracy tend to be correlated and that, if anything, numeracy is lower than literacy; however, as we have noted, many studies do not distinguish these abilities. Research has also not examined the full range of health-related outcomes potentially associated with numeracy. A conspicuous gap in this respect pertains to the relationship between numeracy and patient experiences with care, an outcome domain that is receiving greater attention with the growing emphasis on patient-centered care and communication (R. M. Epstein & Street, 2007). Numeracy might be expected to influence several outcomes in this domain, including patient satisfaction with care, the nature and quality of the patient–physician relationship, and the extent of shared and informed decision making.

Numeracy may have the strongest and most direct connection to the latter outcome of informed decision making, to the extent that this outcome is based on the underlying concept of substantial understanding or gist, for which numeracy represents a critical prerequisite (Reyna & Hamilton, 2001; see Theories of Mathematical Cognition: Psychological Mechanisms of Numeracy, below). For preference-sensitive decisions, that is, decisions involving uncertainty about the net benefits and harms of a medical intervention, some degree of numeracy is necessary for patients to appropriately understand and weigh these benefits and harms (O'Connor, et al., 2007). However, presently there is no empirical evidence demonstrating how numeracy relates to informed decision making or other outcomes in the domain of patient experiences with care.

Although the studies we have reviewed suggest ways in which numeracy may moderate the effects of other psychological factors on different health outcomes, less is known about what factors might moderate the effects of numeracy. The influence of numeracy on health outcomes is likely to be highly context dependent (Montori & Rothman, 2005), varying substantially according to numerous factors that determine the extent and types of numerical reasoning that are required of patients. For example, problems such as patient self-management of anticoagulation therapy clearly require basic arithmetical skills, whereas other problems, such as the interpretation of individualized cancer risk information, involve higher order probabilistic reasoning. However, it is not known how much other routine clinical problems require such skills; in some contexts, numeracy may have no demonstrable effect on health outcomes.

Even when numeracy effects exist, they may be difficult to demonstrate in health outcomes because of the influence of other confounding factors. For example, as noted previously, numeracy was found to have only a modest effect on glycemic control in diabetic patients (Cavanaugh et al., 2008). Although this seems counterintuitive given that self-management of diabetes involves various tasks that are clearly computational in nature (e.g., measuring blood sugar and adjusting insulin and oral medication doses), glycemic control also depends on numerous noncomputational factors, such as diet, weight control, genetic factors, and access

to quality health care. In the context of these other factors, the influence of numeracy may be necessary but not sufficient. In other words, good medical outcomes depend on the occurrence of a series of linked events; any break in the causal chain prior to reaching the outcome can mask essential positive effects at earlier stages.

Potential moderators of numeracy effects include characteristics of the individual and of the health care environment (see Figure 1). Individual differences in personality variables, such as need for cognition, might conceivably moderate numeracy effects by determining the extent to which patients rely on numerical reasoning in the first place, regardless of the extent to which clinical circumstances demand such reasoning. Individuals are known to differ in their preferences for information and participation in health decisions, factors that may further influence the extent to which patients engage in computational tasks (e.g., Reyna & Brainerd, 2008). This engagement may be influenced even more by differences in clinicians' communication and decision-making practices in their encounters with patients.

Research has only begun to identify the important moderators and mediators of the effects of numeracy on health outcomes, and coherent theories that account for the causal mechanisms linking these factors have not been implemented. The critical challenge for future research is not only to identify the unique associations between numeracy—as opposed to literacy—and various proximal and distal health outcomes, but to develop a solid theoretical grounding for this inquiry. A more explicit application of theories of health behavior and decision making to research on numeracy would facilitate identification of a broader range of potentially important moderators and mediators of numeracy effects and the generation of empirically testable hypotheses about the causal mechanisms underlying these effects. We now turn to a discussion of theories that can provide such guidance for future research.

Theories of Mathematical Cognition: Psychological Mechanisms of Numeracy

Four Frameworks

Existing theoretical frameworks make predictions about numeracy, and recent research has begun to exploit these frameworks. There are four major theoretical approaches that are relevant to numeracy: (a) psychophysical approaches in which subjective magnitude is represented internally as a nonlinear function of objective magnitude, (b) computational approaches that stress reducing cognitive load and that emphasize "natural" quantitative processing, (c) dual-process approaches that contrast intuitive (or affective) and analytical processing in which errors are due mainly to imprecision and faulty intuitions, and (d) fuzzy trace theory, a dual-process theory that stresses gist-based intuition as an advanced mode of processing and contrasts it with verbatim-based analytical processing. We cannot review these theories in exhaustive detail, given the scope of the current article, but we outline the approaches, review their implications for numeracy, and point to avenues for future research.

Psychophysical Approaches

Decades of research have shown that humans and animals represent number in terms of language-independent mental magnitudes (Brannon, 2006). The internal representation of number obeys Weber's law, in which the discriminability of two magnitudes is a function of their ratio rather than the difference between them (Gallistel & Gelman, 2005). Thus, the psychological difference between $40 and $20 (a ratio of 2.00) is larger than that between $140 and $120 (a ratio of 1.16), although the absolute difference is identical ($20). If the internal representation were linear, a difference of $20 would always feel like the same amount. The deviation from linearity, in which the same objective difference is not perceived to be identical,

is referred to as a subjective distortion (or an approximate representation) of objective magnitude.

Human infants represent number in this way, showing the same property of ratio-dependent discrimination shared by adult humans and animals (Brannon, 2006). Young children's internal representation of number may be more distorted (or nonlinear) than that of educated adults (Siegler & Opfer, 2003). In addition, Dehaene, Izard, Spelke, and Pica (2008) found that numerical judgments by both children and adults who were members of an indigenous Amazonian group (with a reduced numerical lexicon and little or no formal education) were best fitted with a logarithmic curve, similar to those observed for young children in Western cultures (Siegler & Booth, 2004). In contrast, the responses of adults who had been through a longer educational period were best fitted with a linear function.

The evolutionarily ancient nonverbal numerical estimation system exists alongside a learned verbal system of counting and computational rules. (Number words, such as *two*, can access the nonverbal estimation system, but words are not required in order to process number.) The intraparietal sulcus has been implicated as a substrate for the nonverbal system that represents the meaning of number (Brannon, 2006). The neural correlates of these nonverbal numerical abilities are distinct from those of language-dependent mathematical thinking (Dehaene, Piazza, Pinel, & Cohen, 2003). For example, when adults solve precise, symbolic mathematical problems, their performance is encoded linguistically and engages left inferior frontal regions that are active during verbal tasks. These language areas are not active in neuroimaging studies of nonverbal number processing.

Research in this psychophysical tradition is relevant to explaining causal mechanisms underlying numeracy (e.g., Cantlon & Brannon, 2007; Dehaene, 2007; Furlong & Opfer, 2009; Gallistel & Gelman, 2005; Reyna & Brainerd, 1993, 1994; Shanteau & Troutman, 1992; Siegler & Opfer, 2003). A straightforward implication is that people without brain injury who are low in numeracy retain a nonverbal estimation system for representing number, regardless of their level of formal mathematical knowledge. Moreover, distortions in the perception of numbers should influence judgment and decision making involving numbers (Chen, Lakshminaryanan, & Santos, 2006; Furlong & Opfer, 2009). Indeed, effects reviewed earlier, such as framing, were originally predicted by assuming that the perception of magnitudes (e.g., of money) was distorted in accordance with the psychophysical functions of "prospect theory" (Kahneman & Tversky, 1979).

Building on research on the psychophysics of magnitude, Peters, Slovic, Västfjäll, and Mertz (2008) showed that greater distortions in number perception (i.e., greater deviations from linearity) were associated with lower numeracy, consistent with results for children and less educated adults. They also showed that distortions in number perception predicted greater temporal discounting (choosing a smaller immediate reward over a later larger one) and choosing a smaller proportional (but larger absolute) difference between outcomes (see also Benjamin, Brown, & Shapiro, 2006; Stanovich, 1999). These authors argued that imprecise representations of number magnitudes may be responsible for difficulties in health decision making observed among those low in numeracy.

The Peters et al. (2008) study represents an important first step in linking psychophysical measures of how individuals perceive numbers to measures of health numeracy, and the authors showed that perceptions of number are in turn related to decisions. However, closer inspection of the direction of differences raises questions about what these results signify. All participants tended to choose $100 now rather than wait for $110 in a month, regardless of the precision of their number perception. However, those with more precise representations (i.e., those with more linear representations and thus larger perceived differences between numbers) were more

likely to wait to receive $15 in a week rather than accept $10 now. Ignoring the difference between 4 weeks (a month) and 1 week, those with a more precise representation treated a difference of $5 as though it were larger than a difference of $10. This pattern of preferences is consistent with a ratio-dependent discrimination of number because the ratio of $15 to $10 is larger than the ratio of $110 to $100. Moreover, choosing $15 over $10 but then choosing $100 over $110 could be justified by the different time delays: The $110 is delayed a month, making it less desirable.

The preferences in Peters et al.'s (2008) second experiment are not so easily justified, however. When asked to choose which charitable foundation should receive funding, participants with more precise representations of number were more likely to choose a foundation that reduced deaths from disease from 15,000 a year to 5,000 a year, compared with a foundation that reduced deaths from 160,000 a year to 145,000 a year or even one that reduced deaths from 290,000 a year to 270,000 a year. As in the first experiment, preferences were consistent with a ratio-dependent discrimination of number; a greater proportion of lives saved (67% over 6.9%) was preferred. However, participants with more precise representations of number were more likely to choose the numerically inferior option, to save 5,000 lives rather than save 20,000 (or 15,000) lives. As Peters et al. acknowledge, this choice is not the normatively correct one.

Surprisingly, numeracy was not significantly related to preferences in this task for younger adults (there was a marginal interaction among age, numeracy, and precision of representation, suggesting that higher numeracy and higher precision together increased preference of the inferior option for older adults). If we consider the implications for medical decision making, these results are troubling. People with superior number discrimination would be more likely to choose the worst option in terms of number of lives saved. Unfortunately, this is not an isolated result. As we discuss below, more numerate individuals (who tend to have more precise representations of number) sometimes choose the numerically inferior option in other tasks, too.

Computational Approaches

Computational approaches emphasize reducing cognitive burdens associated with information processing. As an example, working memory limitations are assumed to interfere with cognitive processing, including processing of numerical information. Therefore, reducing memory load (i.e., reducing demands on working memory) is predicted to improve performance (as long as sufficient information is processed for accurate performance; for reviews, see Reyna, 1992, 2005). In this view, poor decision making is the result of information overload and the failure to sufficiently process information, as many have assumed in research on formatting effects reviewed earlier. Improving decision making, then, requires reducing the amount of information to be processed, especially irrelevant information, which drains working memory capacity, while thoroughly processing relevant information (Peters, Dieckmann, et al., 2007).

Consistent with this approach, strategies aimed at making numerical information more organized and accessible have been shown to improve decision making. For example, asking people to actively process information by enumerating reasons for their preferences or by indicating the exact size of a risk on a bar chart is predicted to enhance the use of numbers and reduce reliance on extraneous sources of information (Mazzocco, Peters, Bonini, Slovic, & Cherubini, 2008; Natter & Berry, 2005). In this connection, Mazzocco et al. (2008) found that asking decision makers to give reasons for choices encouraged greater weighting of numerical relative to nonnumerical information (e.g., emotion and anecdotes). Decision analysis and public health programs emphasize this kind of precise and elaborate processing of numerical information (e.g., Fischhoff, 2008). Dual-process theories, discussed below, have incorporated

this computational approach into their assumptions about the analytical side of processing (e.g., S. Epstein, 1994; Peters et al., 2006; Reyna, 2004, 2008; see Nelson et al., 2008).

The natural frequency hypothesis is another computational approach (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1994; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Predictions have not been made about numeracy as differences across individuals; rather, they concern which kinds of numerical displays are more "transparent" than others, given the way that most people process numbers (e.g., Brase, 2002; Cosmides & Tooby, 1996; Gigerenzer, 1994).

As an example of how natural frequencies differ from probabilities, consider the following information: The probability of colorectal cancer is 0.3%. If a person has colorectal cancer, the probability that the Hemoccult test is positive is 50%. If a person does not have colorectal cancer, the probability that he still tests positive is 3%. The same information expressed in terms of natural frequencies would be as follows: Out of every 10,000 people, 30 have colorectal cancer. Of these, 15 will have a positive Hemoccult test. Out of the remaining 9,970 people without colorectal cancer, 300 will still test positive.

Natural frequencies were thought to facilitate reasoning because they reduce the number of required computations. They are "natural" in the sense that they are assumed to correspond to the way in which humans have experienced statistical information over most of their evolutionary history (e.g., Gigerenzer & Hoffrage, 1995).

The hypothesis that frequencies or counts are more natural and easier to process than percentages or decimals (e.g., probabilities) appeared unassailable in the 1990s (e.g., Gigerenzer, Todd, & the ABC Group, 1999). For example, problems framed using natural frequencies were said to elicit fewer biases and errors than problems using probabilities (e.g., Cosmides & Tooby, 1996). However, these predictions have been challenged by a growing body of evidence (for reviews, see Barbey & Sloman, 2007; Reyna & Brainerd, 2008). In particular, the hypothesis that frequencies are easier to understand than probabilities was not confirmed (e.g., Evans, Handley, Perham, Over, & Thompson, 2000; Koehler & Macchi, 2004; Sloman, Over, Slovak, & Stibel, 2003; see also Macchi & Mosconi, 1998). In studies of risk communication and medical decision making, frequency and probability versions of identical information have been compared, and results have also disconfirmed this frequency hypothesis. For example, Cuite, Weinstein, Emmons, and Colditz (2008) studied 16,133 people's performance on multiple computational tasks involving health risks and found that performance was very similar for frequency (55% accurate) and probability (57% accurate) versions of the same information (see also Dieckmann et al., 2009).

Furthermore, biases and heuristics were not reduced by presenting information using frequencies, once confounding factors were eliminated (see Barbey & Sloman, 2007; Reyna & Brainerd, 2008; Reyna & Mills, 2007a). For example, Windschitl (2002) found biasing effects of a context question on subsequent target judgments of cancer risk, but the bias was not less severe when frequency rather than probability representations were used. Complex decisions were also not made easier with frequencies. Waters, Weinstein, Colditz, and Emmons (2006) compared frequencies with percentages to determine which might increase the accuracy of judgments about trade-offs for different cancer treatments. Among 2,601 respondents, those who received the percentages performed significantly better than those who received the identical information in the form of frequencies.

Thus, there is little evidence to support the idea that frequencies per se (when not confounded with other factors) are more natural or easier to comprehend than percentages or other "normalized" formats. However, it should be noted that the claim that all frequencies facilitate judgment should be distinguished from the natural frequencies hypothesis as characterized, for example, by Hoffrage, Gigerenzer, Krauss, and Martignon (2002). First, natural frequencies

only refer to situations in which two variables are involved—they are nonnormalized joint frequencies (e.g., as in the example above of colorectal cancer and Hemoccult test results). Moreover, proponents argue that natural, but not relative, frequencies facilitate judgment. These proposals have much in common with the nested-sets or class-inclusion hypothesis, which holds that overlapping or nested relations create confusion (e.g., Reyna, 1991; Reyna & Mills, 2007a). The natural frequencies format clarifies relations among classes, but frequencies per se appear to be neither a necessary nor a sufficient means of disentangling classes (e.g., Brainerd & Reyna, 1990; Wolfe & Reyna, 2009). The frequency hypothesis of Cosmides, Gigerenzer, and colleagues (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1994) should be distinguished from the frequency effect studied by Slovic and colleagues (discussed below; e.g., Slovic, Finucane, Peters, & MacGregor, 2004).

Dual-Process Approaches

Other theories take a dual-process approach to explain numerical processing (see Gigerenzer & Regier, 1996, for arguments against standard dual-process approaches). Extrapolating from psychodynamic dualism, S. Epstein and colleagues (e.g., see S. Epstein, 1994) have developed a series of measures of analytical or rational thinking versus intuitive or "experiential" thinking. This distinction between analytical and intuitive thought resembles other dual-process distinctions (e.g., Sloman, 1996; Stanovich, 1999) and has been applied by Kahneman (2003), Slovic et al. (2004), and others (e.g., Peters et al., 2006) to account for heuristics and biases. That is, heuristics and biases, which typically violate rules of probability theory or other quantitative rules, are ascribed to a more primitive intuitive way of thinking (System 1) that can sometimes be overridden or censored by advanced analytical thought (System 2).

Some versions of dual-process theory are vulnerable to the criticism that they are, at best, a post hoc typology that does not lend itself to novel prediction, the true hallmark of a scientific theory. However, S. Epstein, Pacini, Denes-Raj, and Heier's (1996) dual-process theory is not post hoc because a valid and reliable instrument has been fashioned to characterize analytical versus intuitive thinking, which can then be used to predict heuristics and biases. As we discuss, however, although the instrument is a satisfactory measure from an empirical standpoint, its predicted relations to heuristics and biases are not obtained consistently. That is, although reliable individual differences in thinking style are detected when using the instrument, the Rational–Experiential Inventory (REI), these differences do not map onto judgments and decision making in ways that this dual-process theory predicts (S. Epstein et al., 1996).

Specifically, the original version of the REI (S. Epstein et al., 1996) consisted of two scales, Need for Cognition and Faith in Intuition, which correspond to analytical and intuitive thinking styles, respectively (drawing on Cacioppo & Petty's, 1982, Need for Cognition scale). The original REI has been improved by adding items, and the reliability of the experiential scale has been increased (Pacini & Epstein, 1999a, 1999b). The new REI retains the good psychometric properties of the old measure, such as producing two orthogonal factors in factor analyses and exhibiting convergent and divergent validity with respect to other aspects of behavior and personality. Thus, the new REI seems to measure what the theory indicates that it ought to measure.

The basic assumption of this dual-process approach as applied to numeracy (e.g., Peters et al., 2006) is that intuitive thinking is the source of biases and errors in numerical (and other) processing. Analytical thinking, in contrast, is the source of accurate and objective numerical processing. Although intuition is not assumed to lead invariably to biases, a key rationale for standard dual-process theory is that systematic biases are caused by intuitive thinking. A core prediction of S. Epstein et al.'s (1996) theory, therefore, is that a predominance of intuitive over analytical thinking, as measured by the REI, will account for an effect that is sometimes called *ratio bias* (the same effect is called the numerosity bias in the probability judgment

literature; for a review, see Reyna & Brainerd, 1994). The ratio (or numerosity) bias is the finding that people who understand that probability is a function of frequencies in both the numerator and the denominator still tend to pay less attention to the denominator as a default.

In the classic ratio bias task derived from Piaget and Inhelder (1951/1975), participants are offered a prize if they draw a colored jelly bean from a bowl. Bowl A contains nine colored and 91 white jelly beans, and Bowl B contains one colored and nine white jelly beans. Consequently, the rational or analytically superior choice is Bowl B: The chance of picking a colored jelly bean is objectively greater if you pick from Bowl B (10% chance of winning) than if you pick from Bowl A (9% chance of winning). The intuitive choice or ratio bias effect, however, is to pick Bowl A because it contains more winning jelly beans than Bowl B (i.e., nine vs. one). The theory predicts that when individual differences favor rational or analytical thought (measured by the Need for Cognition scale), people ought to pick Bowl B. However, those lower in rationality and/or higher in intuition (measured by the Faith in Intuition scale) should exhibit the ratio bias effect by picking Bowl A. Unfortunately, several critical tests of this prediction (including those using the improved REI measure) conducted by the theorists themselves yielded weak and inconsistent results (e.g., Pacini & Epstein, 1999a, 1999b; see also Alonso & Fernández-Berrocal, 2003; Reyna & Brainerd, 2008).

Another example of heuristic processing that has been examined by using this dual-process theory is framing (e.g., Porcelli & Delgado, 2009). Predictions for framing effects are the same as those for the ratio bias, namely, that people high in analytical thinking but low in intuition should be less susceptible to framing effects than those low in analytical thinking and high in intuition. Framing effects occur when decision makers treat quantitatively equivalent options differently, such as rating a person as more intelligent when told that the person received a test score of 80% correct, as opposed to receiving a score of 20% incorrect. Shiloh, Salton, and Sharabi (2002) presented framing problems to college students and analyzed the data using three factors as independent variables: high or low analytical, high or low intuitive, and positive or negative frame. The results showed that participants fitting nonpredicted combinations of thinking styles—high analytical, high intuitive thinking and low analytical, low intuitive thinking—were the only ones to exhibit framing effects. The findings were interpreted as supporting “the individual-differences perspective on heuristic processing, and as a validation of main assumptions” of dual-process theory (Shiloh et al., p. 415). Again, however, core predictions of dual-process theory were not confirmed: Neither low reliance on analysis nor high reliance on intuition was associated in any consistent fashion with framing effects, contrary to the theory.

Despite these null or inconsistent effects, the dual-process theory’s predictions regarding numeracy are straightforward and have met with greater success. According to Peters et al. (2006), for example, those higher in numeracy should approach numerical problems more analytically, whereas those lower in numeracy would be subject to intuitive biases, such as ratio bias and framing effects. In two of four experiments, they found the predicted pattern: Those higher in numeracy were less likely to exhibit ratio bias in one experiment and framing effects in the other experiment. According to Peters et al., the superior results of the more numerate are due to the greater clarity and precision of their perceptions of numbers (see also Peters & Levin, 2008). For instance, high-numerate participants were assumed to select Bowl B because they perceived the objective probabilities more clearly than low-numerate participants.

The results of a third experiment were only partially supportive of dual-process predictions. Consistent with the theory, frequency and percentage formats did not differ for those higher in numeracy, but they differed for those lower in numeracy. In judging the probability of violence, highly numerate people judged 10 out of 100 patients committing a violent act as

equivalent to 10% of patients committing a violent act. However, the low-numerate participants were predicted to rely on affect, considered part of intuition, as opposed to mathematics. According to the theory, relying on affect should lead to higher levels of risk being reported for the frequency (compared with the percentage) format because more vivid images of violent acts are generated (see Peters et al., 2006). Hence, those lower in numeracy should be more susceptible to the emotionally arousing frequency format, compared with those higher in numeracy, who rely on cold numbers. However, inconsistent with this theory, differences between low- and high-numerate participants were observed for the percentage format, not for the frequency format. Both groups seemed to have relied on similar processes in the emotionally arousing frequency condition (Slovic et al., 2004).

In a final experiment, Peters et al. (2006) found that high-numerate participants were more prone than low-numerate participants to an intuitive bias in processing quantitative information. Different groups rated the attractiveness of playing a bet, either 7/36 chances to win $9 and 29/36 chances to win nothing (the no-loss bet) or 7/36 chances to win $9 and 29/36 chances to lose 5 cents (the loss bet). In an earlier study, Slovic et al. (2004) found that the no-loss bet received an average rating of 9.4 on a 21-point desirability scale, but ratings jumped to 14.9 for the loss bet, which added the possible loss of 5 cents. Thus, the objectively worse bet was rated as more attractive. Peters et al. found that those higher in numeracy showed this effect; they gave higher ratings to the objectively worse bet. Those lower in numeracy rated them the same, a result consistent with the fact that the bets are objectively similar.

Peters et al. (2006) acknowledged that rating the worse bet more highly is a less “rational” response. According to Peters et al., the highly numerate may sometimes make less rational responses than the less numerate “precisely because they focus on the detail of numbers” (p. 411). Nevertheless, dual-process theory predicts the opposite, that the highly numerate should show less bias (i.e., their judgments should better reflect objective quantities) than the less numerate. The same theoretical principles that explain the absence of ratio bias and framing effects for the highly numerate appear to be violated when the opposite result, greater bias, is found for the loss bet.

Dual-process theory also predicts that mood will have biasing effects on those low in numeracy. The less numerate, who are less likely to attend to and understand numbers, should be more influenced by extraneous information, such as mood or affect. This effect was demonstrated in a study that examined how people made judgments about hospital quality (Peters et al., in press). Although most numerical quality indicators remained unused by all respondents, the highly numerate were more likely to use one of the indicators to rate the hospitals. As expected, compared with those of high-numerate patients, preferences expressed by low-numerate patients were less influenced by objective probabilities and more influenced by their mood.

In sum, individual differences in dual processes do not consistently predict biases in processing numerical information in ratio bias and framing tasks. Differences in numeracy that are supposed to reflect such dual processes, however, are associated with ratio bias and framing effects as well as with effects of mood. Other effects of numeracy run counter to theoretical predictions: Those higher in numeracy rated a numerically worse bet as superior (those lower in numeracy did not), and numeracy did not produce expected differences in affective processing of numbers in a frequency format. Taken together, these theoretical tests suggest that the hypothesized differences in dual processes do not fully explain effects of numeracy. Future research should be aimed at delineating the specific processes that underlie biases and heuristics in people who differ in numeracy.

Fuzzy Trace Theory

Building on research in psycholinguistics, fuzzy trace theory distinguishes between verbatim and gist representations of information, extending this distinction beyond verbal information to numbers, pictures, graphs, events, and other forms of information (e.g., Reyna & Brainerd, 1992, 1995). Verbatim representations capture the literal facts or "surface form" of information, whereas gist representations capture its meaning or interpretation (based on a person's culture, education, and experience, among other factors known to affect meaning; e.g., Reyna, 2008; Reyna & Adam, 2003). Gist representations are also less precise than verbatim ones; they are the "fuzzy traces" in fuzzy trace theory.

Verbatim and gist representations of information are encoded separately, and each forms the basis for different kinds of reasoning, one focused on memory for precise details (verbatim-based reasoning) and the other on understanding global meaning (gist-based reasoning). Thus, fuzzy trace theory is a dual-process theory but one in which gist-based intuition is an advanced mode of reasoning. Although standard dual-process theories have been criticized for lacking evidence for distinct processes (Keren & Schul, in press), there is extensive evidence for the independence of verbatim and gist processes, including findings from formal mathematical tests (e.g., Brainerd & Reyna, 2005; Reyna & Brainerd, 1995).

Specifically, research has shown that people encode verbatim representations as well as multiple gist representations of the same information. When presented with various numbers or numerosities, people encode verbatim representations of numbers and gist representations that capture the order of magnitudes, whether they are increasing or decreasing (e.g., over time), which magnitudes seem large or small, among other qualitative (inexact) relations (Brainerd & Gordon, 1994; Reyna & Brainerd, 1991a, 1993a, 1994a, 1995; Reyna & Casillas, 2009; see also Gaissmaier & Schooler, 2008). For instance, given quantitative information that the numbers of deaths worldwide are 1.3 million deaths a year for lung cancer, 639,000 for colorectal cancer, and 519,000 for breast cancer, people encode such gists as "lung cancer deaths are most," "lung cancer deaths are more than breast cancer," and so on.

People prefer to operate on the fuzziest or least precise representation that they can use to accomplish a task, such as making a judgment or decision (e.g., Reyna & Brainerd, 1991b, 1994, 1995; Reyna, Lloyd, & Brainerd, 2003). They begin with the simplest distinctions (e.g., categorical) and then move up to more precise representations (e.g., ordinal and interval) as the task demands. For example, fuzzy trace theory accounts for framing effects by assuming that people use the most basic gist for number, the categorical distinction between some quantity and none. Thus, a choice between saving 200 people for sure and a one-third probability of saving 600 people (and two-thirds probability of saving no one) is interpreted as saving some people for sure versus maybe saving some and maybe saving none. Because saving some people is better than saving none, the sure option is preferred. Analogous interpretations of the loss frame (as a choice between some people dying for sure vs. maybe some dying and maybe none dying) produces preferences for the gamble because none dying is better than some dying.

More generally, framing effects occur because numbers are interpreted semantically in terms of vague relations, such as good versus bad, low versus high, some versus none, or more versus less (Mills et al., 2008; Reyna, 2008; Reyna et al., 2003). Often these gist interpretations reflect affect (see Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008; Rivers, Reyna, & Mills, 2008). The specific explanation for framing effects described above has been confirmed by experiments (e.g., Kühberger & Tanner, 2009; Reyna & Brainerd, 1991b, 1995). Psychophysical accounts of framing are not sufficient to explain the results of these experiments. For example, framing effects persist even when some or all of the numbers in framing problems are deleted, and contrary to psychophysical predictions, framing effects are

actually larger under these circumstances. Conversely, focusing attention on the numbers that are supposed to generate framing effects (in the psychophysical accounts) makes the effects smaller.

As people gain experience making certain judgments or decisions, they tend to rely more on gist rather than verbatim representations, known as a *fuzzy-processing preference* (e.g., Nelson et al., 2008; Reyna & Ellis, 1994; Reyna & Lloyd, 2006). For example, framing effects have been predicted and found to increase from childhood to adulthood (Reyna, 1996; Reyna & Ellis, 1994); other heuristics and biases show a similar, counterintuitive trend (see Reyna & Farley, 2006, Table 3). In adulthood, experts have been found to base their decisions more on simple gist, compared with novices with less experience and knowledge (e.g., Reyna & Lloyd, 2006). Relying on gist may be especially beneficial for older people whose verbatim memories are less robust (Reyna & Mills, 2007b; Tanius, Wood, Hanoch, & Rice, 2009). Age differences in choice quality between younger and older adults are reduced when decisions are based on affect or bottom-line valence (Mikels et al., in press; but see Peters et al., 2008).

These and other developmental trends suggest that more advanced numerical processing is not necessarily more precise or elaborate, as assumed in standard dual-process theories (e.g., Peters et al., 2006), but rather that it involves the extraction of bottom-line meanings or relations. Hence, knowing the best estimate of a lifetime risk of dying from breast cancer (such as that provided by online calculators) or knowing the exact probability of complications from surgery does not constitute informed consent or informed medical decision making, according to fuzzy trace theory (e.g., Reyna, 2008; Reyna & Hamilton, 2001). People can receive a precise estimate of risk tailored for them and yet not understand the essential gist of whether their risk is low or high, whether they should be relieved or alarmed.

In fact, focusing on exact numbers has been found to exacerbate some biases. Removing numerical information so that participants must rely, instead, on memory for its gist improves performance (e.g., in class-inclusion problems; see Brainerd & Reyna, 1995; Reyna & Brainerd, 1995). The ratio bias effect is an example of a class-inclusion illusion; assumptions about retrieval and processing, as well as about representations, are required to explain this effect (Reyna, 1991; Reyna & Mills, 2007a; Wolfe & Reyna, 2009). Briefly, any ratio concept, including probability, is inherently confusing because the referents of classes overlap. Owing to this confusion, people focus on the target classes in numerators (e.g., the nine colored jelly beans in Bowl A and the one colored jelly bean in Bowl B) and neglect the classes in the denominator (e.g., the 100 total jelly beans in Bowl A and the 10 total jelly beans in Bowl B), producing the ratio bias effect (Reyna & Brainerd, 1994, Figure 11.3). Like the participants who favored saving proportionately more lives in Peters et al.'s (2008) experiment, people who favor Bowl A are making comparisons of relative magnitude, but they are the wrong comparisons (Reyna, 1991; Reyna & Brainerd, 2008). Experiments manipulating the salience of the wrong relative magnitude—a competing gist—confirm that this factor contributes to the ratio bias effect (e.g., Brainerd & Reyna, 1990, 1995).

As fuzzy trace theory also predicts, manipulations that reduce confusion by discretely representing classes or drawing attention to denominators can reduce the ratio bias effect (e.g., Brainerd & Reyna, 1990, 1995; F. J. Lloyd & Reyna, 2001; Wolfe & Reyna, 2009). Because processing problems due to overlapping classes are not fundamental logical errors (i.e., participants understand the role of numerators and denominators in principle), class-inclusion errors persist even for advanced reasoners (see also Chapman & Liu, 2009). For example, physicians and high school students performed equally poorly on a base-rate neglect problem, which is another type of class-inclusion problem (Reyna, 2004; Reyna & Adam, 2003). They estimated the probability of disease for a patient with a positive test result, given the base rate of disease, and confused that positive predictive value with the test's sensitivity (probability

of a positive test result for a patient with disease), as expected by fuzzy theory, because only their denominators differ (Reyna & Mills, 2007a). Therefore, icon arrays or other formats that disentangle classes (clarifying their relations), especially those that also make the relevant gist salient, reduce these biases.

In sum, fuzzy trace theory distinguishes intuitive reasoning based on simple gist representations from detailed quantitative analysis using verbatim representations (a distinction supported by much independent evidence). The theory can account for heuristics and biases that have been the foci of earlier theories, such as framing and ratio bias effects. Although earlier approaches emphasized precision, accuracy, and analysis, fuzzy trace theory holds that more advanced reasoners (i.e., those who understand the meaning of numbers) should rely on the simple, qualitative gist of numbers whenever the task permits. Although research on numeracy is consistent with fuzzy trace theory, especially with its core assumption of fuzzy processing as a default mode of reasoning, alternative theoretical explanations for specific findings have rarely been compared (cf. Reyna & Brainerd, 2008). Current health numeracy measures seem to capture the ease or automaticity with which ratios are computed, an explanation that would account broadly for performance across problems for which the computed ratios were and were not appropriate. Tests of numeracy have been not yet been devised that capture individual differences in appropriate gist processing of numbers.

Overview and Future Directions

We began this article with a description of the dilemma of low health numeracy. Despite the abundance of health information from commercial and noncommercial sources, including information about major new research discoveries that can be used to prevent and treat disease, most people cannot take advantage of this abundance. Few problems can be said to affect up to 93 million people, based on reliable assessments of nationally representative samples. Low numeracy is such a problem. The ideal of informed patient choice, in which patients share decision making with health care providers, is an elusive goal without the ability to understand numerical information about survival rates, risks of treatments, and conditional probabilities that govern such domains as genetic risk (e.g., the probability of disease given a genetic mutation). Those who are disadvantaged by poverty, lack of education, or linguistic barriers are also unlikely to have numerical skills that would empower them to access health care and to make informed decisions.

Definitions of health numeracy—encompassing computational, analytical, and statistical skills, among other abilities—are impressively broad, and yet, on assessments of all varieties, people cannot accomplish much less ambitious tasks, such as judging whether a .001 risk of death is bigger or smaller than 1 in 100. Moreover, scores on numeracy assessments have been linked to key cognitions that predict morbidity and mortality, to health behaviors, and, in a few cases, to medical outcomes, the latter sometimes only indirectly through measures of literacy that include, and are correlated with, numeracy. Evidence of effects of numeracy exists along a causal chain from initial perceptions of risks and benefits to health-related judgments and decisions, which have been found to be biased and inaccurate for people with low numeracy. Low numeracy has been shown to impair understanding of risks and benefits of cancer screening, to reduce medication compliance in anticoagulation therapy, to limit access to preventive treatments for asthma, and to affect known predictors of death and disability, such as patients' self-rated functional status.

However, there are many gaps and shortcomings in current research on health numeracy. The health domains that have been studied (e.g., breast cancer risk perception and screening) have been limited. For example, despite its importance, we could find no research on the effects of numeracy in mental health (e.g., on medication compliance in treatment for depression).

Research has documented strong effects of numeracy on perceptions of risks and benefits, on elicitation of values or utilities, and on formatting effects such as framing and frequency effects, but only a handful of studies connect such perceptions, values, and effects to health behaviors or outcomes. Finally, and most important, much of the work is merely descriptive, rather than explanatory or, as scientific theory ought to be, predictive based on knowledge of causal mechanisms.

Although evocative and practical, none of the definitions of numeracy is based on empirically informed, theoretically sound conceptions of numeracy. Assessments are similarly pragmatic rather than explanatory, despite evidence of their "validity" and reliability. On the basis of studies that have controlled for education, intelligence, literacy, and other factors, we can be reasonably sure that numeracy is a separate faculty. What that faculty consists of is the province of theory. Several theorists have characterized it as the ability to draw meaning from numbers, although they disagree about whether that meaning is affective, frequentistic, precisely quantitative, or fuzzy gist. The idea that people vary in the quality of the meaning that they extract from numbers is central to characterizing them as low or high in numeracy. Clearly, more sophisticated and coherent conceptual definitions and measures of numeracy are needed to account for the diverse, sometimes inconsistent ways in which numeracy has been found to relate to decision making and other outcomes.

The pervasive theme that those low in numeracy score lower on just about every other dimension studied makes sense, and it is consistent with dual-process theories that contrast intuitive and analytical reasoning and attribute biases and fallacies mainly to the former. However, these theories do not explain surprising and robust exceptions to this rule, including nonnumerical framing effects, inconsistent relations between intuitive versus analytical thinking and biases, and greater preference for numerically inferior options (e.g., saving fewer rather than more lives or a loss bet over a no-loss bet) among those higher in numeracy. Furthermore, standard dual-process theories emphasize affect, which fails to account for some effects, such as frequency, but is implicated in others, such as mood. The surprising findings generated by dual-process theories are informative precisely because they challenge conventional assumptions about numeracy, precision, and accurate reasoning. These anomalies should be a focus of future research in order to better understand the mechanisms of numerical processing.

Each of the theories we reviewed has been applied to pitfalls in numerical processing or to heuristics and biases. Psychophysical approaches fall short in this respect. They explain the ratio dependence of number perception, which can influence decision making involving numbers, but they do not explain ratio bias. This is a serious shortcoming because ratio concepts —fractions, decimals, percentages, and probabilities—are especially difficult to process, as observed in national and international surveys, as well as in many kinds of numeracy assessments. This difficulty is expected, according to fuzzy trace theory, because class-inclusion judgments of all kinds (e.g., in logical reasoning and in judgments of nested probabilities, such as 5-year vs. lifetime risks of cancer) are subject to denominator neglect, explaining ratio bias, frequency effects, and confusion of conditional probabilities, among other findings. The theory also identifies specific interventions to reduce denominator neglect, which have been evaluated with populations ranging from children to physicians and been found effective. Contemporary theory seems to be coalescing around the conclusion that computational simplicity—that is, clarifying relations among classes—is important for understanding. However, little work on individual differences in numeracy has been done from a computational perspective.

Although many of the most important questions for future research on numeracy have implications for theory, some questions do not hinge on any particular theoretical perspective,

such as how to better distinguish numeracy from automatic computation or general reasoning ability. However, the most informative research would test specific hypotheses about how people who are low versus high in numeracy process information differently. Are specific results produced by affect or gist, by frequencies or denominator neglect, and what kinds of meaning do highly numerate people extract from important health information? Most imperative, how can such meaning be communicated more broadly to those who need it to make life-and-death decisions?

Without a deeper theoretical understanding of numeracy, especially of deficiencies in numeracy, it is difficult to know which policy recommendations to make. However, one important question raised by the association between numeracy and outcomes is whether clinical screening for low numeracy should be implemented in health care settings. The data that we have reviewed suggest the potential utility of numeracy screening as a means of helping clinicians to identify low-numerate patients at risk for poor understanding of health information and to avert more distal adverse health outcomes through interventions targeted to these patients.

For a number of reasons, however, the prospect of clinical screening for low numeracy is not straightforward. As we have noted, the evidence linking low numeracy and poor health outcomes is newly emerging and much less developed than the evidence on health literacy. There is currently no evidence that either numeracy screening or targeted interventions to improve numeracy or otherwise assist low-numerate patients will improve health outcomes. Although it stands to reason that this should be the case, one can argue that more evidence is needed before such a practice is implemented, particularly given the substantial resources that numeracy screening would likely entail in the clinical setting. Some researchers have advanced the same argument regarding clinical health literacy screening, which also lacks direct empirical support in spite of the larger evidence base linking health literacy and outcomes (Paasche-Orlow & Wolf, 2008).

Other important considerations in assessing the prospect of clinical screening for low numeracy include the performance characteristics of the screening tests, and the potential harms of numeracy screening. Currently, there are several tools that could be used to screen for low health numeracy, although none has been widely accepted or validated for clinical purposes. Screening for low numeracy also has unknown acceptability and psychological effects on patients' experiences with health care, and these factors require further exploration before screening programs are implemented. Although similar concerns have been expressed about health literacy screening, limited evidence suggests that patients have favorable attitudes toward screening (Ryan et al., 2007); more work needs to be done to determine whether these findings generalize to health numeracy.

A larger question relates to the optimal approach of the health care system to the problem of low numeracy. Clinical screening for low health numeracy represents an individual-based approach, aimed at detecting the risk factor of low numeracy and, in theory, targeting interventions toward high-risk individuals. An alternative population-based approach, however, would be to design communication and care interventions that would benefit all patients, regardless of their individual numeracy levels. For example, clinical interventions to improve the understandability of numerical information and to evaluate and ensure comprehension of this information might benefit all patients, even those with high numeracy. Supporting this possibility, research on health literacy suggests that educational interventions designed to target low-literacy individuals also benefit those with high literacy (DeWalt et al., 2006). If this is also true for numeracy, then one can ask whether the more worthwhile strategy would be to implement more broadly applicable interventions to improve numerical understanding. These approaches, however, are not mutually exclusive, and the optimal

strategy is an empirical question. Once sufficient evidence is gathered, it may be feasible to add effectiveness in overcoming innumeracy as a quality indicator in the evaluation of procedures used in hospitals (e.g., for surgical consent) and in clinical practice.

Acknowledgments

This research was supported in part by National Cancer Institute Grant R13CA126359, National Science Foundation Grant BCS-0840111, and National Institute of Mental Health Grant RO1-MH-061211 to Valerie F. Reyna. We would also like to acknowledge the assistance of Kathryn Hambleton.

References

Adelswärd V, Sachs L. The meaning of 6.8: Numeracy and normality in health information talks. Social Science and Medicine 1996;43:1179–1187. [PubMed: 8903121]

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. [PubMed: 10022112]

Alonso D, Fernández-Berrocal P. Irrational decisions: Attending to numbers rather than ratios. Personality and Individual Differences 2003;35:1537–1547.

Ancker JS, Kaufman D. Rethinking health numeracy: A multidisciplinary literature review. Journal of the American Medical Informatics Association 2007;14:713–721. [PubMed: 17712082]

Apter AJ, Cheng J, Small D, Bennett IM, Albert C, Fein DG, Van Horne S. Asthma numeracy skill and health literacy. Journal of Asthma 2006;43(9):705–710. [PubMed: 17092853]

Apter AJ, Paasche-Orlow MK, Remillard JT, Bennett IM, Ben-Joseph EP, Batista RM, Rudd RE. Numeracy and communication with patients: They are counting on us. Journal of General Internal Medicine 2008;23:2117–2124. [PubMed: 18830764]

Artinian, NT.; Lange, MP.; Templin, TN.; Stallwood, LG.; Hermann, CE. Functional health literacy in an urban primary care clinic. Internet Journal of Advanced Nursing Practice. 2003. Retrieved from http://www.ispub.com/journal/the_internet_journal_of_advanced_nursing_practice.html

Baker DW. The meaning and the measure of health literacy. Journal of General Internal Medicine 2006;21 (8):878–883. [PubMed: 16881951]

Baker DW, Gazmararian JA, Sudano J, Patterson M. The association between age and health literacy among elderly persons. Journals of Gerontology: Series B: Psychological Sciences and Social Sciences 2000;55:S368–S374.

Baker DW, Gazmararian JA, Sudano J, Patterson M, Parker RM, Williams MV. Health literacy and performance on the Mini-Mental State Examination. Aging & Mental Health 2002;6:22–29. [PubMed: 11827619]

Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, Peel J. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. American Journal Public Health 2002;92(8):1278–1283.

Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, Peel J. Health literacy and use of outpatient physician services by Medicare managed care enrollees. Journal of General Internal Medicine 2004;19(3):215–220. [PubMed: 15009775]

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. [PubMed: 9844076]

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(6):1027–1030. [PubMed: 9224190]

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(1):33–42. [PubMed: 14528569]

Barbey AK, Sloman SA. Base-rate respect: From ecological rationality to dual processes. Behavioral and Brain Sciences 2007;30:241–254. [PubMed: 17963533]

Bass PF III, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. Academic Medicine 2002;77:1039–1041. [PubMed: 12377684]

Benjamin, DJ.; Brown, SA.; Shapiro, JM. Who is "behavioral"? Cognitive ability and anomalous preferences. 2006. Unpublished manuscript. Retrieved from http://www.arts.cornell.edu/econ/dbenjamin/Who_is_Behavioral_Cognitive_ability_and_anomalous_preferences.pdf

Benson JG, Forman WB. Comprehension of written health care information in an affluent geriatric retirement community: Use of the Test of Functional Health Literacy. Gerontology 2002;48:93–97. [PubMed: 11867931]

Berger CR. Base-rate bingo: Ephemeral effects of population data on cognitive responses, apprehension, and perceived risk. Communication Research 2002;29(2):99–124.

Bernhardt, JM.; Brownfield, ED.; Parker, RM. Understanding health literacy. In: Schwartzberg, JG.; VanGeest, JB.; Wang, CC., editors. Understanding health literacy: Implications for medicine and public health. Chicago, IL: American Medical Association Press; 2005. p. 3-16.

Black WC, Nease RF, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. Journal of the National Cancer Institute 1995;87:720–731. [PubMed: 7563149]

Bogardus ST Jr, Holmboe E, Jekel JF. Perils, pitfalls, and possibilities in talking about medical risk. Journal of the American Medical Association 1999;281:1037–1041. [PubMed: 10086441]

Brainerd CJ, Gordon LL. Development of verbatim and gist memory for numbers. Developmental Psychology 1994;30:163–177.

Brainerd CJ, Reyna VF. Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. Developmental Review 1990;10:365–403.

Brainerd CJ, Reyna VF. Autosuggestibility in memory development. Cognitive Psychology 1995;28:65–101. [PubMed: 7895469]

Brainerd, CJ.; Reyna, VF. The science of false memory. New York, NY: Oxford University Press; 2005.

Brainerd CJ, Stein LM, Silveira RA, Rohenkohl G, Reyna VF. How does negative emotion cause false memories? Psychological Science 2008;19:919–925. [PubMed: 18947358]

Brannon EM. The representation of numerical magnitude. Current Opinion in Neurobiology 2006;16:222–229. [PubMed: 16546373]

Brase GL. Ecological and evolutionary validity: Comments on Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni's (1999) mental-model theory of extensional reasoning uncertainty. Psychological Review 2002;109:722–728. [PubMed: 12374326]

Brewer NT, Weinstein ND, Cuite CL, Herrington J. Risk perceptions and their relation to risk behavior. Annals of Behavioral Medicine 2004;27:125–130. [PubMed: 15026296]

Buchbinder R, Hall S, Youd JM. Functional health literacy of patients with rheumatoid arthritis attending a community-based rheumatology practice. Journal of Rheumatology 2006;33:879–886. [PubMed: 16511936]

Burkell J. What are the chances? Evaluating risk and benefit information in consumer health materials. Journal of the Medical Library Association 2004;92(2):200–208. [PubMed: 15098049]

Cacioppo JT, Petty RE. The need for cognition. Journal of Personality and Social Psychology 1982;42:116–131.

Cantlon JF, Brannon EM. Adding up the effects of cultural experience on the brain. Trends in Cognitive Sciences 2007;11(1):1–4. [PubMed: 17129750]

Cavanaugh K, Huizinga MM, Wallston KA, Gebretsadik T, Shintani A, Davis D, Rothman RL. Association of numeracy and diabetes control. Annals of Internal Medicine 2008;148(10):737–746. [PubMed: 18490687]

Chapman GB, Liu J. Numeracy, frequency, and Bayesian reasoning. Judgment and Decision Making 2009;4:34–40.

Chen MK, Lakshminaryanan V, Santos LR. The evolution of our preferences: Evidence from capuchin monkey trading behavior. Journal of Political Economy 2006;114(3):517–537.

Cockroft, WH. Mathematics counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the chairmanship of W. H. Cockcroft. London, England: Her Majesty's

Stationery Office; 1982. Retrieved from http://www.dg.dial.pipex.com/documents/docs1/cockcroft02.shtml

Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition 1996;58:1–73.

Cuite CL, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Medical Decision Making 2008;28:377–384. [PubMed: 18480036]

Davids SL, Schapira MM, McAuliffe TL, Nattinger AB. Predictors of pessimistic breast cancer risk perceptions in a primary care population. Journal of General Internal Medicine 2004;19(4):310–315. [PubMed: 15061739]

Davis T, Crouch M, Long S, Jackson R, Bates P, George R, Bairnsfather L. Rapid assessment of literacy levels of adult primary care patients. Family Medicine 1991;6:433–435. [PubMed: 1936717]

Davis, T.; Kennen, EM.; Gazmararian, JA.; Williams, MV. Literacy testing in health care research. In: Schwartzberg, JG.; VanGeest, JB.; Wang, CC., editors. Understanding health literacy: Implications for medicine and public health. Chicago, IL: American Medical Association Press; 2005. p. 157-179.

Dehaene, S. Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In: Haggard, P.; Rossetti, Y.; Kawato, M., editors. Attention and Performance: Vol. 22. Sensorimotor foundations of higher cognition. Cambridge, MA: Harvard University Press; 2007. p. 527-574.

Dehaene S, Izard V, Spelke E, Pica P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. Science 2008 May 30;320:1217–1220. [PubMed: 18511690]

Dehaene S, Piazza M, Pinel P, Cohen L. Three parietal circuits for number processing. Cognitive Neuropsychology 2003;20(3–6):487–506.

DeWalt DA, Malone RM, Bryant ME, Kosnar MC, Corr KE, Rothman RL, Pignone MP. A heart failure self-management program for patients of all literacy levels: A randomized, controlled trial. BMC Health Services Research 2006;6:30.10.1186/1472-6963-6-30 [PubMed: 16533388]

Dieckmann NF, Slovic P, Peters E. The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. Risk Analysis. Advance online publication. 200910.1111/j.1539-6924.2009.01279.x

Dillard AJ, McCaul KD, Kelso PD, Klein WMP. Resisting good news: Reactions to breast cancer risk communication. Health Communication 2006;19(2):115–123. [PubMed: 16548702]

Doak LG, Doak CC. Patient comprehension profiles: Recent findings and strategies. Patient Counseling and Health Education 1980;2:101–106.

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.

Educational Testing Service. Test content for Health Activities Literacy Tests. 2007. Retrieved from http://www.ets.org/etsliteracy

Epstein, RM.; Street, RL. Patient-centered communication in cancer care: Promoting healing and reducing suffering (NIH Report No. 07-6225). Bethesda, MD: National Cancer Institute; 2007.

Epstein S. Integration of the cognitive and psychodynamic unconscious. American Psychologist 1994;49:709–724. [PubMed: 8092614]

Epstein S, Pacini R, Denes-Raj V, Heier H. Individual differences in intuitive–experiential and analytical–rational thinking styles. Journal of Personality and Social Psychology 1996;71:390–405. [PubMed: 8765488]

Estrada C, Barnes V, Collins C, Byrd JC. Health literacy and numeracy. Journal of the American Medical Association 1999;282(6):527. [PubMed: 10450711]

Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC. Literacy and numeracy skills and anticoagulation control. American Journal of the Medical Sciences 2004;328(2):88–93. [PubMed: 15311167]

Evans, JStBT.; Handley, SJ.; Perham, N.; Over, DE.; Thompson, VA. Frequency versus probability formats in statistical word problems. Cognition 2000;77:197–213. [PubMed: 11018509]

Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: Present and future research in risk communication. American Journal of Health Behavior 2007;31(Suppl 1):S47–S56. [PubMed: 17931136]

Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. Medical Decision Making 2007;27(5):672–680. [PubMed: 17641137]

Fischhoff B. Assessing adolescent decision-making competence. Developmental Review 2008;28:12–28.

Forrow L, Taylor WC, Arnold RM. Absolutely relative: How research results are summarized can affect treatment decisions. American Journal of Medicine 1992;92(2):121–124. [PubMed: 1543193]

Furlong E, Opfer JE. Cognitive constraints on how economic rewards affect cooperation. Psychological Science 2009;20:11–16. [PubMed: 19037906]

Gaissmaier W, Schooler LJ. The smart potential behind probability matching. Cognition 2008;109:416–422. [PubMed: 19019351]

Gal, I. Big picture: What does “numeracy” mean?. GED Items. 1995. Retrieved from http://mathforum.org/teachers/adult.ed/articles/gal.html

Gallistel, CR.; Gelman, R. Mathematical cognition. In: Holyoak, KJ.; Morrison, RG., editors. The Cambridge handbook of thinking and reasoning. New York, NY: Cambridge University Press; 2005. p. 559-588.

Gazmararian JA, Baker DW, Williams MV, Parker RM, Scott TL, Green DC, Koplan JP. Health literacy among Medicare enrollees in a managed care organization. Journal of the American Medical Association 1999;281:545–551. [PubMed: 10022111]

Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases. Journal of General Internal Medicine 2006;21:1215–1221. [PubMed: 17105519]

Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Education and Counseling 2003;51(3):267–275. [PubMed: 14630383]

Gigerenzer, G. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: Wright, G.; Ayton, P., editors. Subjective probability. New York, NY: Wiley; 1994. p. 129-161.

Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychological Science in the Public Interest 2007;8:53–96.

Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review 1995;102:684–704.

Gigerenzer G, Regier T. How do we tell an association from a rule? Comment on Sloman (1996). Psychological Bulletin 1996;119:23–26.

Gigerenzer, G.; Todd, PM. the ABC Research Group. Simple heuristics that make us smart. New York, NY: Oxford University Press; 1999.

Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke S. A definition and operational framework for health numeracy. American Journal of Preventive Medicine 2005;29(4):375–376. [PubMed: 16242604]

Gong DA, Lee JY, Rozier RG, Pahel BT, Richman JA, Vann WF Jr. Development and testing of the Test of Functional Health Literacy in Dentistry (TOFHLiD). Journal of Public Health Dentistry 2007;67:105–112. [PubMed: 17557682]

Gonzales, P.; Guzmán, JC.; Partelow, L.; Pahlke, E.; Jocelyn, L.; Kastberg, D.; Williams, T. Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003 (NCES Report No. 2005-005). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2004.

Gramling R, Irvin JE, Nash J, Sciamanna C, Culpepper L. Numeracy and medicine: Key family physician attitudes about communicating probability with patients. Journal of the American Board of Family Medicine 2004;17(6):473.

Greene J, Peters E, Mertz CK, Hibbard JH. Comprehension and choice of a consumer-directed health plan: An experimental study. American Journal of Managed Care 2008;14(6):369–376. [PubMed: 18554075]

Grigg, W.; Donahue, P.; Dion, G. The nation’s report card: 12th-grade reading and mathematics 2005 (NCES Report No. 2007-468). Washington, DC: U.S. Department of Education, National Center for

Education Statistics; 2007. Retrieved from http://nationsreportcard.gov/reading_math_grade12_2005/s0301.asp

Gurmankin AD, Baron J, Armstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. Medical Decision Making 2004a;24(3):265–271. [PubMed: 15155015]

Gurmankin AD, Baron J, Armstrong K. Intended message versus message received in hypothetical physician risk communications: Exploring the gap. Risk Analysis 2004b;24(5):1337–1347. [PubMed: 15563299]

Haggstrom DA, Schapira MM. Black–White differences in risk perceptions of breast cancer survival and screening mammography benefit. Journal of General Internal Medicine 2006;21(4):371–377. [PubMed: 16686816]

Hibbard JH, Peters E. Supporting informed consumer health care decisions: Data presentation approaches that facilitate the use of information in choice. Annual Review of Public Health 2003;24:413–433.

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. [PubMed: 17684108]

Hibbard JH, Slovic P, Peters E, Finucane ML, Tusler M. Is the informed-choice policy approach appropriate for Medicare beneficiaries? Health Affairs 2001;20(3):199–203. [PubMed: 11585167]

Hoffrage U, Gigerenzer G, Krauss S, Martignon L. Representation facilitates reasoning: What natural frequencies are and what they are not. Cognition 2002;84:343–352. [PubMed: 12044739]

Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. Science 2000 December 22;290:2261–2262. [PubMed: 11188724]

Huizinga MM, Elasy TA, Wallston KA, Cavanaugh K, Davis D, Gregory RP, Rothman RL. Development and validation of the Diabetes Numeracy Test (DNT). BMC Health Services Research 2008;8:96.10.1186/1472-6963-8-96 [PubMed: 18452617]

Institute of Medicine. Health literacy: A prescription to end confusion. Washington, DC: National Academies Press; 2004.

Jastak, S.; Wilkinson, GS. Wide Range Achievement Test: Administration manual. Wilmington, DE: Jastak Assessment Systems; 1984. rev

Johnston MV, Diab ME, Kim SS, Kirshblum S. Health literacy, morbidity, and quality of life among individuals with spinal cord injury. Journal of Spinal Cord Medicine 2005;28:230–240. [PubMed: 16048141]

Joint Committee on National Health Education Standards. National health education standards: Achieving health literacy. Atlanta, GA: American Cancer Society; 1995.

Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist 2003;58:697–720. [PubMed: 14584987]

Kahneman D, Tversky A. Prospect theory: An analysis of decisions under risk. Econometrica 1979;47:313–327.

Kalichman SC, Ramachandran B, Catz S. Adherence to combination antiretroviral therapies in HIV patients of low health literacy. Journal of General Internal Medicine 1999;14:267–273. [PubMed: 10337035]

Keren G, Schul Y. Two is not always better than one: A critical evaluation of two-system theories. Perspectives on Psychological Science. in press.

Kicklighter JR, Stein MA. Factors influencing diabetic clients' ability to read and comprehend printed diabetic diet material. The Diabetes Educator 1993;19:40–46. [PubMed: 8458298]

Kirsch, IS.; Jungeblut, A.; Jenkins, L.; Kolstad, A. Adult literacy in America: A first look at the findings of the National Adult Literacy Survey (NCES Report No. 1993-275. 3. Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2002.

Klein WMP, Stefanek ME. Cancer risk elicitation and communication: Lessons from the psychology of risk perception. CA: A Cancer Journal for Clinicians 2007;57:147–167. [PubMed: 17507441]

Koehler J, Macchi L. Thinking about low-probability events: An exemplar cuing theory. Psychological Science 2004;15:540–546. [PubMed: 15270999]

Kühberger A. The influence of framing on risky decisions: A meta-analysis. Organizational Behavior and Human Decision Processes 1998;75(1):23–55. [PubMed: 9719656]

Kühberger A, Tanner C. Risky choice framing: Task versions and a comparison of prospect theory and fuzzy-trace theory. Journal of Behavioral Decision Making. Advance online publication. 200910.1002/bdm.656

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 Report No. 2006-483). Washington, DC: Government Printing Office; 2006.

Lee, J.; Grigg, W.; Dion, G. The nation's report card: Mathematics 2007 (NCES Report No. 2007-494). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2007. Retrieved from http://nces.ed.gov/nationsreportcard/mathematics

Lemke, M.; Sen, A.; Pahlke, E.; Partelow, L.; Miller, D.; Williams, T.; Jocelyn, L. International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective (NCES Report No. 2005-003). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2004.

Lenert LA, Sherbourne CD, Reyna VF. Utility elicitation using single item questions compared with a computerized interview. Medical Decision Making 2001;21:97–104. [PubMed: 11310952]

Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P. 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. [PubMed: 12015518]

Lipkus IM, Hollands JG. The visual communication of risk. Journal of the National Cancer Institute Monographs 1999;25:149–163. [PubMed: 10854471]

Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical Decision Making 2001;21:37–44. [PubMed: 11206945]

Lloyd FJ, Reyna VF. A web exercise in evidence-based medicine using cognitive theory. Journal of General Internal Medicine 2001;16:94–99. [PubMed: 11251760]

Lloyd, G. Ministry of Education, 15 to 18: A report of the Central Advisory Council for Education (England). Vol. 1. London, England: Her Majesty's Stationery Office; 1959.

Macchi L, Mosconi G. Computational features vs. frequentist phrasing in the base-rate fallacy. Swiss Journal of Psychology 1998;57(2):79–85.

Maibach E. Cancer risk communication: What we know and what we need to learn. Journal of the National Cancer Institute Monographs 1999;25:179–181. [PubMed: 10854478]

Marcus EN. The silent epidemic—The health effects of illiteracy. New England Journal of Medicine 2006;355:339–341. [PubMed: 16870912]

Mayben JK, Kramer JR, Kallen MA, Franzini L, Lairson DR, Giordano TP. Predictors of delayed HIV diagnosis in a recently diagnosed cohort. AIDS Patient Care 2007;21:195–204.

Mazzocco K, Peters E, Bonini N, Slovic P, Cherubini P. How priming analytical versus emotional thinking influences choice processes. 2008 Unpublished manuscript.

McNeal B, Salisbury Z, Baumgardner P, Wheeler FC. Comprehension assessment of diabetes education program participants. Diabetes Care 1984;7:232–235. [PubMed: 6734392]

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. [PubMed: 7070445]

Mikels JA, Loeckenhoff C, Maglio S, Goldstein M, Garber A, Carstensen LL. Follow your heart or your head? The effects of affective versus deliberative decision strategies differ by age group. Journal of Experimental Psychology: Applied. in press.

Mills BA, Reyna VF, Estrada S. Explaining contradictory relations between risk perception and risk taking. Psychological Science 2008;5:429–434. [PubMed: 18466401]

Montalto NJ, Spiegler GE. Functional health literacy in adults in a rural community health center. West Virginia Medical Journal 2001;97:111–114. [PubMed: 11392190]

Montori V, Leung T, Thompson C, Chung J, Capes S, Smith S. Validation of a diabetes numeracy evaluation tool. Diabetes 2004;53(Suppl 2):A224–A225.

Montori VM, Rothman RL. Weakness in numbers: The challenge of numeracy in health care. Journal of General Internal Medicine 2005;20(11):1071–1072. [PubMed: 16307637]

National Mathematics Advisory Panel. Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education; 2008.

Natter HM, Berry DC. Effects of active information processing on the understanding of risk information. Applied Cognitive Psychology 2005;19:123–135.

Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: Theory and practice. Annals of Behavioral Medicine 2008;35(3):261–274. [PubMed: 18677452]

Nurss JR, El-Kebbi IM, Gallina DL, Ziemer DC, Musey VC, Lewis S, Phillips LS. Diabetes in urban African Americans: Functional health literacy of municipal hospital outpatients with diabetes. The Diabetes Educator 1997;23:563–568. [PubMed: 9355373]

O'Connor AM, Wennberg JE, Legare F, Llewellyn-Thomas HA, Moulton BW, Sepucha KR, King JS. Toward the "tipping point": Decision aids and informed patient choice. Health Affairs (Millwood) 2007;26:716–725.

Paasche-Orlow MK, Wolf MS. Evidence does not support clinical screening of literacy. Journal of General Internal Medicine 2008;23:100–102. [PubMed: 17992564]

Pacini R, Epstein S. The interaction of three facets of concrete thinking in a game of chance. Thinking and Reasoning 1999a;5:303–325.

Pacini R, Epstein S. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. Journal of Personality and Social Psychology 1999b; 76:972–987. [PubMed: 10402681]

Parikh N, Parker R, Nurss J, Baker D, Williams M. Shame and health literacy: The unspoken connection. Patient Education and Counseling 1996;27:33–39. [PubMed: 8788747]

Parker RM, Baker DW, Williams MV, Nurss JR. The Test of Functional Health Literacy in Adults: A new instrument for measuring patients' literacy skills. Journal of General Internal Medicine 1995;10 (10):537–541. [PubMed: 8576769]

Parrott R, Silk K, Dorgan K, Condit C, Harris T. Risk comprehension and judgments of statistical evidentiary appeals: When a picture is not worth a thousand words. Human Communication Research 2005;31(3):423–452.

Paulos, JA. Innumeracy: Mathematical illiteracy and its consequences. New York, NY: Hill and Wang; 1988.

Perie, M.; Grigg, WS.; Dion, GS. The nation's report card: Mathematics 2005 (NCES Report No. 2006-453). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2005.

Perie, M.; Moran, R.; Lutkus, AD. NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics (NCES Report No. 2005-464). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2005.

Peters E, Dieckmann NF, Dixon A, Hibbard JH, Mertz CK, Slovic P. Less is more in presenting quality information to consumers. Medical Care Research and Review 2007;64:169–190. [PubMed: 17406019]

Peters E, Dieckmann NF, Västfjäll D, Mertz CK, Slovic P, Hibbard JH. Bringing meaning to numbers: The impact of evaluative categories on decisions. Journal of Experimental Psychology: Applied. in press.

Peters E, Hibbard JH, Slovic P, Dieckmann NF. Numeracy skill and the communication, comprehension, and use of risk and benefit information. Health Affairs 2007;26(3):741–748. [PubMed: 17485752]

Peters E, Levin IP. Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. Judgment and Decision Making 2008;3:435–448.

Peters E, Slovic P, Västfjäll D, Mertz C. Intuitive numbers guide decisions. Judgment and Decision Making 2008;3:619–635.

Peters E, Västfjäll D, Slovic P, Mertz C, Mazzocco K, Dickert S. Numeracy and decision making. Psychological Science 2006;17(5):407–413. [PubMed: 16683928]

Piaget, J.; Infelder, B. The origin of the idea of chance in children. Leake, L., Jr; Burrell, P.; Fishbein, HD., translators. New York, NY: Norton; 1975.

Politi MC, Han PKJ, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. Medical Decision Making 2007;27:681–695. [PubMed: 17873256]

Porcelli AJ, Delgado MR. Acute stress modulates risk taking in financial decision making. Psychological Science 2009;20:278–283. [PubMed: 19207694]

Ratzan, SC.; Parker, RM. Health literacy (Current Bibliographies in Medicine No. 2000-1). Bethesda, MD: National Library of Medicine; 2000. Retrieved from http://www.nlm.nih.gov/archive/20061214/pubs/cbm/hliteracy.html

Reed AE, Mikels JA, Simon KI. Older adults prefer less choice than young adults. Psychology and Aging 2008;23:671–675. [PubMed: 18808256]

Reyna VF. Class inclusion, the conjunction fallacy, and other cognitive illusions. Developmental Review 1991;11:317–336.

Reyna, VF. Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In: Howe, ML.; Brainerd, CJ.; Reyna, VF., editors. Development of long-term retention. New York, NY: Springer-Verlag; 1992. p. 103-127.

Reyna VF. Conceptions of memory development with implications for reasoning and decision making. Annals of Child Development 1996;12:87–118.

Reyna VF. How people make decisions that involve risk: A dual-processes approach. Current Directions in Psychological Science 2004;13:60–66.

Reyna, VF. Fuzzy-trace theory, judgment, and decision-making: A dual-processes approach. In: Izawa, C.; Ohta, N., editors. Human learning and memory: Advances in theory and application: The 4th Tsukuba International Conference on Memory. Mahwah, NJ: Erlbaum; 2005. p. 239-256.

Reyna VF. A theory of medical decision making and health: Fuzzy-trace theory. Medical Decision Making 2008;28:829–833. [PubMed: 19020341]

Reyna VF, Adam MB. Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. Risk Analysis 2003;23:325–342. [PubMed: 12731817]

Reyna VF, Brainerd CJ. Fuzzy-trace theory and children's acquisition of mathematical and scientific concepts. Learning and Individual Differences 1991a;3:27–59.

Reyna VF, Brainerd CJ. Fuzzy-trace theory and framing effects in choice. Gist extraction, truncation, and conversion. Journal of Behavioral Decision Making 1991b;4:249–262.

Reyna, VF.; Brainerd, CJ. A fuzzy-trace theory of reasoning and remembering: Paradoxes, patterns, and parallelism. In: Hearst, N.; Kosslyn, S.; Shiffrin, R., editors. Essays in honor of William K. Estes: Vol. 2. From learning processes to cognitive processes. Hillsdale, NJ: Erlbaum; 1992. p. 235-259.

Reyna, VF.; Brainerd, CJ. Fuzzy memory and mathematics in the classroom. In: Davies, GM.; Logie, RH., editors. Memory in everyday life. Amsterdam, the Netherlands: North-Holland; 1993. p. 91-119.

Reyna, VF.; Brainerd, CJ. The origins of probability judgment: A review of data and theories. In: Wright, G.; Ayton, P., editors. Subjective probability. New York, NY: Wiley; 1994. p. 239-272.

Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences 1995;7(1):1–75.

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.

Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learning and Individual Differences 2008;18:89–107.

Reyna, VF.; Casillas, W. Development and dual processes in moral reasoning: A fuzzy-trace theory approach. In: Ross, BH.; Bartels, DM.; Bauman, CW.; Skitka, LJ.; Medin, DL., editors. Psychology of Learning and Motivation: Vol. 50. Moral judgment and decision making. San Diego, CA: Academic Press; 2009. p. 207-236.

Reyna VF, Ellis SC. Fuzzy-trace theory and framing effects in children's risky decision making. Psychological Science 1994;5(5):275–279.

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):1–44.

Reyna VF, Hamilton AJ. The importance of memory in informed consent for surgical risk. Medical Decision Making 2001;21:152–155. [PubMed: 11310949]

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. [PubMed: 16953744]

Reyna, VF.; Lloyd, FJ.; Brainerd, CJ. Memory, development, and rationality: An integrative theory of judgment and decision making. In: Schneider, S.; Shanteau, J., editors. Emerging perspectives on judgment and decision research. New York, NY: Cambridge University Press; 2003. p. 201-245.

Reyna VF, Lloyd FJ, Whalen P. Genetic testing and medical decision making. Archives of Internal Medicine 2001;161:2406–2408. [PubMed: 11700152]

Reyna VF, Mills BA. Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). Behavioral and Brain Sciences 2007a;30:278–280.

Reyna, VF.; Mills, BA. Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development. In: Gorfein, DS.; MacLeod, CM., editors. Inhibition in cognition. Washington, DC: American Psychological Association; 2007b. p. 185-210.

Rivers SE, Reyna VF, Mills B. Risk taking under the influence: A fuzzy-trace theory of emotion in adolescents. Developmental Review 2008;28:107–144. [PubMed: 19255597]

Rogers ES, Wallace LS, Weiss BD. Misperception of medical understanding in low-literacy patients: Implications for cancer prevention. Cancer Control 2006;13:225–229. [PubMed: 16885919]

Rosenthal R. The "file drawer problem" and tolerance for null results. Psychological Bulletin 1979;86:638–641.

Rothman RL, Housam R, Weiss H, Davis D, Gregory R, Gebretsadik T, Elasy TA. Patient understanding of food labels: The role of literacy and numeracy. American Journal of Preventive Medicine 2006;31:391–398. [PubMed: 17046410]

Rothman RL, Montori VM, Cherrington A, Pigone MP. Perspective: The role of numeracy in healthcare. Journal of Health Communication 2008;13:583–595. [PubMed: 18726814]

Rudd, RE.; Colton, T.; Schacht, R. An overview of medical and public health literature addressing literacy issues: An annotated bibliography (NCSALL Report No. 14). Cambridge, MA: National Center for the Study of Adult Learning and Literacy; 2000.

Ryan JG, Leguen F, Weiss BD, Albury S, Jennings T, Velez F, Salibi N. Will patients agree to have their literacy skills assessed in clinical practice? Health Education Research 2007;23:603–611. [PubMed: 17890757]

Safeer RS, Keenan J. Health literacy: The gap between physicians and patients. American Family Physician 2005;72:463–468. [PubMed: 16100861]

Schapira MM, Davids SL, McAuliffe TL, Nattinger AB. Agreement between scales in the measurement of breast cancer risk perceptions. Risk Analysis 2004;24(3):665–673. [PubMed: 15209937]

Schapira MM, Fletcher KE, Gilligan MA, King TK, Laud PW, Matthews BA, Hayes E. A framework for health numeracy: How patients use quantitative skills in health care. Journal of Health Communication 2008;13:501–517. [PubMed: 18661390]

Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Bindman AB. Association of health literacy with diabetes outcomes. Journal of the American Medical Association 2002;288:475–482. [PubMed: 12132978]

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(11):966–972. [PubMed: 9412301]

Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the Medical Data Interpretation Test. Medical Decision Making 2005;25(3):290–300. [PubMed: 15951456]

Schwartz SR, McDowell J, Yueh B. Numeracy and the shortcomings of utility assessment in head and neck cancer patients. Head & Neck 2004;26(5):401–407. [PubMed: 15122656]

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(5):395–404. [PubMed: 11961474]

Sentell TL, Halpin HA. Importance of adult literacy in understanding health disparities. Journal of General Internal Medicine 2006;21:862–866. [PubMed: 16881948]

Shanteau J, Troutman CM. A psychophysical evaluation of diminishing returns in riskless decision making. Organizational Behavior and Human Decision Processes 1992;52:569–579.

Sheridan SL, Harris RP, Woolf SH. Shared decision making about screening and chemoprevention: A suggested approach from the U.S. Preventive Services Task Force. American Journal of Preventive Medicine 2004;26:56–66. [PubMed: 14700714]

Sheridan SL, Pignone M. Numeracy and the medical students' ability to interpret data. Effective Clinical Practice 2002;5(1):35–40. [PubMed: 11874195]

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(11):884–892. [PubMed: 14687273]

Shiloh S, Salton E, Sharabi D. Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. Personality and Individual Differences 2002;32:415–429.

Shuchman M. Drug risks and free speech—Can Congress ban consumer drug ads? New England Journal of Medicine 2007;356(22):2236–2239. [PubMed: 17476002]

Siegler RS, Booth JL. Development of numerical estimation in young children. Child Development 2004;75:428–444. [PubMed: 15056197]

Siegler RS, Opfer J. The development of numerical estimation: Evidence for multiple representations of numerical quantity. Psychological Science 2003;14(3):237–243. [PubMed: 12741747]

Sloman SA. The empirical case for two systems of reasoning. Psychological Bulletin 1996;119:3–22.

Sloman SA, Over DE, Slovak L, Stibel JM. Frequency illusions and other fallacies. Organizational Behavior and Human Decision Processes 2003;91:296–309.

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. [PubMed: 15078302]

Smerecnik CM, Mesters I. Validating the Medical Data Interpretation Test in a Dutch population. Patient Education and Counseling 2007;68:287–290. [PubMed: 17686603]

Stanovich, KE. Who is rational? Studies of individual differences in reasoning. Mahwah, NJ: Erlbaum; 1999.

Stanovich KE, West RF. Individual differences in framing and conjunction effects. Thinking and Reasoning 1998;4:289–317.

Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability. Journal of Personality and Social Psychology 2008;94:672–695. [PubMed: 18361678]

Tanius BE, Wood S, Hanoch Y, Rice T. Aging and choice: Applications to Medicare Part D. Judgment and Decision Making 2009;4:92–101.

U.S. Department of Health and Human Services. Healthy people 2010: Understanding and improving health. 2. Vol. 1. Washington, DC: Government Printing Office; 2000. Retrieved from http://www.healthypeople.gov/document/HTML/Volume1/11HealthCom.htm

Waters EA, Weinstein ND, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. Journal of Health Communication 2006;11:167–182. [PubMed: 16537286]

Weinfurt KP, Castel LD, Li Y, Sulmasy DP, Balshem AM, Benson AB 3rd, Meropol NJ. The correlation between patient characteristics and expectations of benefit from Phase I clinical trials. Cancer 2003;98(1):166–175. [PubMed: 12833469]

Weinfurt KP, DePuy V, Castel LD, Sulmasy DP, Schulman KA, Meropol NJ. Understanding of an aggregate probability statement by patients who are offered participation in Phase I clinical trials. Cancer 2005;103(1):140–147. [PubMed: 15534885]

Weinstein ND. What does it mean to understand a risk? Dimensions of risk comprehension. Journal of the National Cancer Institute Monographs 1999;25:15–20. [PubMed: 10854451]

Weiss, BD. Health literacy: A manual for clinicians. Chicago, IL: American Medical Association Foundation and American Medical Association; 2003. Retrieved from http://www.ihconline.org/toolkits/HealthLiteracy/AMAHealthLiteracyManualClinicians.pdf

Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, Hale FA. Quick assessment of literacy in primary care: The newest vital sign. Annals of Family Medicine 2005;3:514–522. [PubMed: 16338915]

Wilkinson, GS. Wide Range Achievement Test 3. Wilmington, DE: Wide Range; 1993.

Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A. Inadequate literacy is a barrier to asthma knowledge and self-care. Chest 1998;114(4):1008–1015. [PubMed: 9792569]

Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease: A study of patients with hypertension and diabetes. Archives of Internal Medicine 1998;158(2):166–172. [PubMed: 9448555]

Williams MV, Parker RM, Baker DW, Parikh NS, Pitkin K, Coates WC, Nurss JR. Inadequate functional health literacy among patients at two public hospitals. Journal of the American Medical Association 1995;274:1677–1682. [PubMed: 7474271]

Windschitl P. Judging the accuracy of a likelihood judgment: The case of smoking risk. Journal of Behavioral Decision Making 2002;15:19–35.

Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Archives of Internal Medicine 2005;165:1946–1952. [PubMed: 16186463]

Wolfe CR, Reyna VF. Semantic coherence and fallacies in estimating joint probabilities. Journal of Behavioral Decision Making. Advance online publication. 200910.1002/bdm.650

Woloshin S, Schwartz LM, Black WC, Welch HG. Women's perceptions of breast cancer risk: How you ask matters. Medical Decision Making 1999;19:221–229. [PubMed: 10424829]

Woloshin S, Schwartz LM, Moncur M, Gabriel S, Tosteson ANA. Assessing values for health: Numeracy matters. Medical Decision Making 2001;21(5):382–390. [PubMed: 11575488]

Woloshin S, Schwartz LM, Welch HG. The value of benefit data in direct-to-consumer drug ads. Health Affairs 2004;23(3):W4-234–W4-245. [PubMed: 15452006]

Woloshin S, Schwartz LM, Welch HG. Patients and medical statistics: Interest, confidence, and ability. Journal of General Internal Medicine 2005;20(11):996–1000. [PubMed: 16307623]

Woloshin S, Schwartz LM, Welch HG. The effectiveness of a primer to help people understand risk: Two randomized trials in distinct populations. Annals of Internal Medicine 2007;146:256–265. [PubMed: 17310049]

World Health Organization. Health promotion glossary. Geneva, Switzerland: Author; 1998. Retrieved from http://www.who.int/hpr/NPH/docs/hp_glossary_en.pdf

Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale (SNS): Effects of low numeracy on comprehension of risk communications and utility elicitations. Medical Decision Making 2007;27:663–671. [PubMed: 17652180]

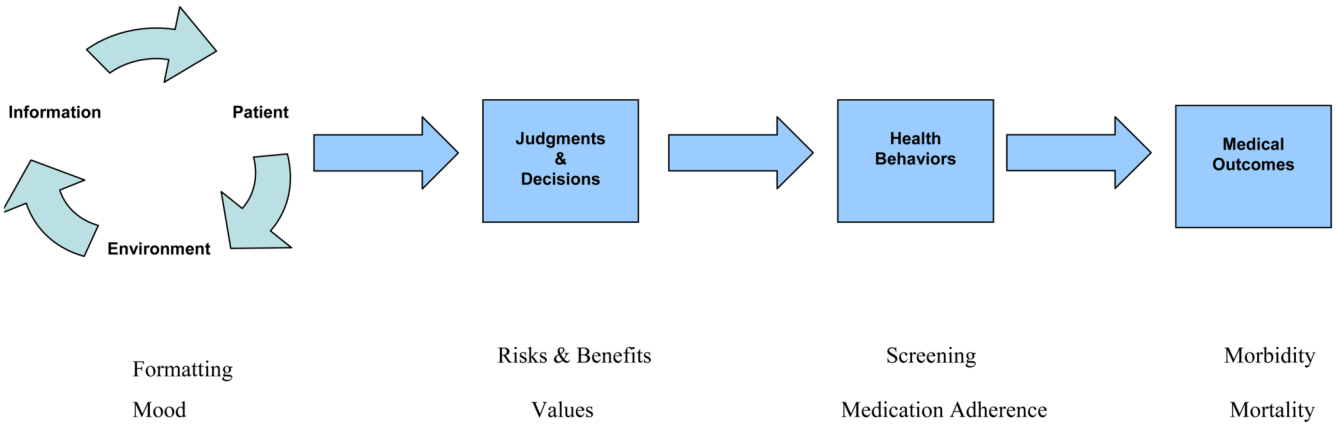

Figure 1.

Table 1

Definitions of Numeracy, Health Numeracy, and Health Literacy

Term	Definition	Source
Numeracy	A word to represent the mirror image of literacy	G. Lloyd (1959, para. 398)
Numeracy	We would wish the word “numerate” to imply the possession of two attributes. The first of these is an “at-homeness” with numbers and an ability to make use of mathematical skills which enables an individual to cope with the practical mathematical demands of his everyday life. The second is an ability to have some appreciation and understanding of information which is presented in mathematical terms, for instance in graphs, charts or tables or by reference to percentage increase or decrease.	Cockroft (1982, para. 34)
Numeracy	The term numeracy describes the aggregate of skills, knowledge, beliefs, dispositions, and habits of mind—as well as the general communicative and problem-solving skills—that people need in order to effectively handle real-world situations or interpretative tasks with embedded mathematical or quantifiable elements.	Gal (1995, para. 9)
Numeracy	Numeracy, in the sense of knowledge and mastery of systems for quantification, measurement and calculation, is a practice-driven competence rather than abstract academic knowledge of “mathematics.” Proficiency in numeracy varies with people’s backgrounds and experience.	Adelswärd and Sachs (1996, p. 1186)
Numeracy	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 mathematic concepts need to be applied when interpreting specific situations.	Montori and Rothman (2005, p. 1071)
Quantitative literacy	The knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed materials.	Kirsch et al. (2002, pp. 3–4)
Health literacy	The capacity of individuals to obtain, interpret and understand basic health information and services and the competence to use such information and services in ways which are health-enhancing.	Joint Committee on National Health Education Standards (1995, p. 5)
Health literacy	A constellation of skills, including the ability to perform basic reading and numerical tasks required to function in the health care environment. Patients with adequate health literacy can read, understand, and act on health care information.	Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs (1999, p. 553)
Health literacy	[Those] cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health…. Health literacy implies the achievement of a level of knowledge, personal skills and confidence to take action to improve personal and community health by changing personal *lifestyles* and *living conditions*.	World Health Organization (1998. p. 10)
Health literacy	The degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions.	Ratzan and Parker (2000, p. vi) [a]
Health numeracy	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.	Golbeck et al. (2005, p. 375)

Term	Definition	Source
Health numeracy	The individual-level skills needed to understand and use quantitative health information, including basic computation skills, ability to use information in documents and non-text formats such as graphs, and ability to communicate orally.	Ancker and Kaufman, D. (2007, p. 713)

[a]Definition also used by the U.S. Department of Health and Human Services (2000, Terminology section) and Institute of Medicine (2004, p. 2).

Table 2

Summary of Measures Used to Assess Health Numeracy

Measures[a]	Reliability	Validity	Test–retest	Reading level (grade)	Components	Administration time	Administration method/response format	Original study population	Correlates
Composite measures									
TOFHLA (Parker et al., 1995)	α = .981; Spearman–Brown = .92	Correlated with REALM: .84 ($p < .001$); correlated with WRAT-R (reading): .74 ($p < .001$)	NS	Reading items: 4.3–19.5; numeracy items: 9.4	50 reading comprehension and 17 numeracy questions based on actual hospital materials	Up to 22 min	Reading: paper and pencil/multiple choice; numeracy: subject views prompts and verbally responds to interviewer questions/ open response	Public hospital patients; predominantly African American, low income, less than high school education	Physical health outcomes, mental health outcomes, [b] disease knowledge, health services utilization
TOFHLA–Spanish (Parker et al., 1995)	α = .98; Spearman–Brown = .84	See Note.	NS	Reading items: 4.3–19.5; numeracy items: 9.4	50 reading comprehension and 17 numeracy questions based on actual hospital materials	Up to 22 min	Same as TOFHLA–English	Public hospital patients; predominantly Hispanic, low income, less than high school education	Same as TOFHLA–English
TOFHLiD (Gong et al., 2007)	α (overall) = .63; α (reading) = .65; α (numeracy) = .59	Correlated with REALD-99: .82 ($p < .05$); TOFHLA: .52 ($p < .05$); REALM: .53 ($p < .05$)	NS	Reading items: 7.0–17.0; numeracy items: 11	68 reading comprehension and 12 numeracy questions based on actual patient materials	20–30 min[c]	Same as TOFHLA	Caretakers of pediatric dental patients; predominantly female, Caucasian, well educated; wide range of income	Child's oral health status
Integrative measures									
Newest Vital Sign (Weiss et al., 2005)	α = .76	Correlated with TOFHLA: .59 ($p < .001$); REALM: .41 ($p < .001$)	NS	NS	Nutrition label accompanied by 6 questions involving arithmetical computation and problem solving	About 3 min	Subject views nutrition label and verbally responds to interviewer questions/open response and multiple choice	University-affiliated primary-care patients; mainly Hispanic and Caucasian; mean years of education = 12.7	NS
Newest Vital Sign–Spanish (Weiss et al., 2005)	α = .69	Correlated with TOFHLA–Spanish: .49 ($p < .001$)	NS	NS	Nutrition label accompanied by 6 questions involving arithmetical computation and problem solving	About 3.5 min	Same as Newest Vital Sign–English	Hispanic university-affiliated primary-care patients; mean years of education = 10.7	NS
Nutrition Label Survey (Rothman et al., 2006)	*KR20* = .87	Correlated with WRAT-3 (math): .67 ($p < .001$); REALM: .47 ($p < .001$)[c]	NS	8–10[c]	24 questions pertaining to interpreting actual food labels; involves arithmetical computation and problem solving	About 30 min	Subject views nutrition label and verbally responds to interviewer questions/open response and multiple choice	University-affiliated primary-care patients; predominantly Caucasian; socioeconomically and educationally diverse	Income, education, literacy, numeracy

Measures[a]	Reliability	Validity	Test–retest	Reading level (grade)	Components	Administration time	Administration method/response format	Original study population	Correlates
Numeracy Assessment (L. M. Schwartz et al., 1997)	α = .56–.80[c]	NS	*r* = .72[c]	7[c]	3 numeracy questions pertaining to probability and converting metrics (percentages and proportions)	2 min[c]	Paper and pencil/open response	Older female veterans; predominantly Caucasian and low income; most with more than high school education	Risk perception, utility assessment
Numeracy Scale (Lipkus et al., 2001)	α = .70–.75	NS	NS	NS	11 numeracy questions covering orders of magnitude, probability, converting metrics, arithmetical computation	Up to 30 min	Paper and pencil/open response and multiple choice	Community volunteers; predominantly Caucasian with more than high school education	See Peters, Dieckmann, et al. (2007) numeracy scale
Expanded Numeracy Scale (Peters, Dieckmann, et al., 2007)	α = .83	High numerate comprehended more and made more high-quality choices in one of two studies; correlated with S-TOFHLA: .51.	NS	NS	15 numeracy questions covering risk magnitude, probability, converting metrics, arithmetical computation	6–15 min[c]	Paper and pencil/open response and multiple choice	Community volunteers; predominantly Caucasian and low income; wide range of educational attainment	Education, self-rated health, SF-8 physical, SF-8 mental, health literacy
Medical Data Interpretation Test (L. M. Schwartz et al., 2005)	α = .71	High-numerate, high-literate, better educated people scored higher than people lower on those characteristics; physician experts performed better than postgraduates	*r* = .67	NS	20 multiple-choice questions designed to test the ability to compare risks and put risk estimates in context	NS	Paper and pencil/multiple choice	Volunteers from the community and local hospitals; wide range of income and educational attainment	NS
Numeracy Assessment (anticoagulation control; Estrada et al., 2004)	NS	NS	NS	NS	6 numeracy questions pertaining to probability, converting metrics (percentages and proportions), arithmetical computation, problem solving	NS	Paper and pencil/open response and multiple choice	Patients from anticoagulation management clinics; predominantly Caucasian; 21% with less than 8th-grade education	Health outcomes (anticoagulation control)
Asthma Numeracy Questionnaire (Apter et al., 2006)	α = .57	Correlated with S-TOFHLA: .47; REALM: .41	NS	NS	4 numeracy questions covering arithmetical computation,	About 5 min[c]	Subject views questionnaire while interviewer reads aloud/open response and multiple choice	Asthma patients from urban clinics; largely female and African American; wide range of educational attainment	Hospitalization, emergency room visits

Measures[a]	Reliability	Validity	Test–retest	Reading level (grade)	Components	Administration time	Administration method/response format	Original study population	Correlates
					percentages, problem solving				
Diabetes Numeracy Test (Huizinga et al., 2008)	*KR20* = .95	Correlated with REALM: .54 ($p < .001$); WRAT (math): .62 ($p < .001$); Diabetes Knowledge Test: .71 ($p < .001$)	NS	6.5[c]	43 items pertaining to arithmetical operations, ratios, multistep mathematics, time, counting, hierarchy	20–46 min	Paper and pencil/open response	Convenience sample of diabetic patients from academic and Veterans Administration health centers; most with less than 9th-grade numeracy	Income, education, insulin adjustment for carbohydrates, insulin adjustment for blood glucose
Diabetes Numeracy Test–15[d] (Huizinga et al., 2008)	*KR20* = .90; *KR20* = .89	Correlated with Diabetes Numeracy Test: .96, .97; REALM: .54, .52; WRAT (math): .65, .61; Diabetes Knowledge Test: .73, .67	NS	6[c]	15 items pertaining to nutrition, exercise, blood glucose monitoring, oral medication, insulin administration	10–15 min	Paper and pencil/open response	Convenience sample of diabetic patients from academic and Veterans Administration health centers; most with less than 9th-grade numeracy	Income, education, insulin adjustment for carbohydrates;[c] insulin adjustment for blood glucose[c]
Subjective Numeracy Scale (Fagerlin, Zikmund-Fisher, et al., 2007)	α = .82	Correlated with Lipkus et al. (2001) numeracy scale: .63–.68	NS	NS	8 questions covering perceived mathematical ability and preference for presentation of statistical information	About 5 min	Paper and pencil/multiple choice	Volunteers recruited from local hospitals; predominantly Caucasian; wide range of educational attainment	NS
STAT–Interest Scale (Woloshin et al., 2005)	α = .70	Correlated with Medical Data Interpretation Test: .26 ($p = .006$)	$r = .60$	NS	4 statements and 1 question pertaining to interest in medical statistics	NS	Paper and pencil/multiple choice	Volunteers from the community and local hospitals; predominantly Caucasian; wide range of income and educational attainment	Intention to use statistics for health decision making
STAT–Confidence Scale (Woloshin et al., 2005)	α = .78	Correlated with Medical Data Interpretation Test: .15 ($p = .04$)	$r = .62$	NS	2 statements and 1 question pertaining to confidence understanding medical statistics	NS	Paper and pencil/multiple choice	Volunteers from the community and local hospitals; predominantly Caucasian; wide range of income and educational attainment; multiple choice	NS

Note. For the Test of Functional Health Literacy in Adults (TOFHLA) and TOFHLA–Spanish, reliability and validity were calculated for the entire TOFHLA, not for the individual reading comprehension and numeracy sections. Validity of the TOFHLA–Spanish could not be determined because the Rapid Estimate of Adult Literacy in Medicine (REALM) and Wide Range Achievement Test–Revised (WRAT-R) are not available in Spanish. NS = not specified; TOFHLiD = Test of Functional Health Literacy in Dentistry; REALD-99 = Rapid Estimate of Adult Literacy in Dentistry; WRAT-3 = WRAT (3rd ed.; Wilkinson, 1993); S-TOFHLA = short version of the TOFHLA; SF-8 = Short Form (8) Health Survey.

[a]The Health Activities Literacy Tests (Baker, 2006; Educational Testing Service [ETS], 2007) are not included among the numeracy measures in this table because they are proprietary and available only commercially from ETS. However, they bear mentioning because they assess the ability of 16- to 65-year-old adults to understand and act upon health-related written materials. Like the National Adult Literacy Survey, these tests include prose, document, and quantitative literacy items that reflect practical health-related tasks. They address health activities related to health promotion and protection, disease prevention, health care and maintenance, and systems navigation. The Health Activities Literacy Tests are only available electronically. What is unique about the tests is that they adapt to the test taker's ability and are scored by

computer. The full-length Health Activities Literacy Test requires one hour to complete, whereas the shorter “Locator” test takes about 30–40 min to complete. Currently, these tests are used primarily for research purposes (http://www.ets.org/Media/Tests/ETS_Literacy/LITERACY_TECHSPECS.pdf).

[b] Study used original S-TOFHLA.

[c] Personal communication from instrument developer.

[d] The Diabetes Numeracy Test–15 was verified through random split sample analysis; reliability and validity data for the test are given for both subsamples.