7-2013

Health Numeracy: The Importance of Domain in Assessing Numeracy

Helen Levy
University of Michigan - Ann Arbor, hlevy@umich.edu

Peter A. Ubel

Amanda J. Dillard
Grand Valley State University, dillaram@gvsu.edu

David R. Weir

Angela Fagerlin

Health Numeracy: The Importance of Domain in Assessing Numeracy

Helen Levy, PhD, Peter A. Ubel, MD, Amanda J. Dillard, PhD, David R. Weir, PhD, Angela Fagerlin, PhD

Background and Objective. *Existing research concludes that measures of general numeracy can be used to predict individuals' ability to assess health risks. We posit that the domain in which questions are posed affects the ability to perform mathematical tasks, raising the possibility of a separate construct of "health numeracy" that is distinct from general numeracy. The objective was to determine whether older adults' ability to perform simple math depends on domain.* ***Methods.*** *Community-based participants completed 4 math questions posed in 3 different domains: a health domain, a financial domain, and a pure math domain. Participants were 962 individuals aged 55 and older, representative of the community-dwelling US population over age 54.* ***Results.*** *We found that respondents performed significantly worse when questions were posed in the health domain (54% correct) than in either the pure math domain (66% correct) or the financial domain (63% correct). Our experimental measure of numeracy consisted of only 4 questions, and it is possible that the apparent effect of domain is specific to the mathematical tasks that these questions require.* ***Conclusions.*** *These results suggest that health numeracy is strongly related to general numeracy but that the 2 constructs may not be the same. Further research is needed into how different aspects of general numeracy and health numeracy translate into actual medical decisions.* ***Key words:*** *health numeracy; health literacy; cognition.* ***(Med Decis Making XXXX;XX:XXX–XXX)***

A growing literature documents the impact of numeracy—"the ability to comprehend, use, and attach meaning to numbers"[1]—on medical decision making.[2] Individuals with low numeracy, compared with those who have higher numeracy, are less likely to understand health risk or to comply with medication regimes,[3,4] are known to underuse screening for colorectal cancer,[5] have greater difficulty managing chronic conditions,[6,7] and report worse subjective health.[8] The mechanisms through which low numeracy translates into worse medical decision making and health remain active areas for research. What is clear, however, is that low numeracy is widespread. Most people perform poorly on numeracy tests; according to the 2003 National Assessment of Adult Literacy, only about 13% of adults were proficient in "quantitative literacy."[9] Even highly educated individuals have difficulty with fairly simple math problems.[10]

Studies of numeracy and medical decision making have relied on a range of measures. Numeracy is measured using both objective measures such as math tests[10,11] and subjective measures such as individuals' own assessments of their quantitative ability.[12,13] Numeracy may also be assessed using problems that are purely mathematical, or in a way that is specific to health and/or medical care or even specific to a particular disease such as diabetes or asthma.[14–16] An interesting set of unanswered questions about the measurement of numeracy as it relates to medical decision making concerns the role of domain. Does

Received 11 April 2012 from Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (HL, DRW); School of Public Health, University of Michigan, Ann Arbor, MI (HL); Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, MI (HL); Fuqua School of Business, Sanford School of Public Policy, and School of Medicine, Duke University, Durham, NC (PAU); Department of Psychology, Grand Valley State University, Allendale, MI (AJD); VA Ann Arbor Center for Clinical Management Research, Ann Arbor, MI (AF); Division of General Internal Medicine, University of Michigan, Ann Arbor, MI (AF); and Center for Bioethics and Social Sciences in Medicine, Ann Arbor, MI (AF). Financial support: The Health and Retirement Study is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. Levy acknowledges financial support from the National Institute on Aging (grant number NIA K01AG034232). Revision accepted for publication 15 May 2013.

Address correspondence to Helen Levy, Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48104; e-mail: hlevy@umich.edu.

DOI: 10.1177/0272989X13493144

domain matter—that is, does quantitative ability depend on whether questions are situated in specific health domains (''10% of 1000 patients . . .'') versus financial domains (''10% of $1000 . . .'') versus more general domains (''10% of 1000 . . .'')? Certainly, there is evidence that situating a task in a relevant domain can enhance performance, such as Cosmides and Tooby's classic demonstration that reasoning in the Wason card sort task is enhanced when it is presented in a contextualized scenario (carding drinkers) relative to the abstract ''pure reasoning'' version.[17] If this principle extends to mathematical proficiency, with individuals showing an increased or decreased ability to solve mathematical problems when presented in a health domain, then this may be evidence that general numeracy and health numeracy are separate constructs.

Golbeck and others (2005) proposed a distinct concept of ''health numeracy,'' but very little research has explored the distinction between numeracy and health numeracy, either conceptually or empirically.[18–20] One of the few empirical studies of health numeracy versus general numeracy was conducted by Lipkus and colleagues.[10] Those authors recruited participants through newspaper advertisements to participate in 4 separate studies pertaining to breast and colon cancer screening. Each study had between 121 and 126 participants; combining all 4 yields a sample of 463 participants, aged 40 and older, approximately four-fifths of whom were women. Participants completed a general numeracy questionnaire consisting of basic mathematical questions similar to those used by Schwartz and others[11] and then completed an expanded numeracy questionnaire that posed similar questions in terms of health (e.g., the probability of developing a disease).[10]

The central finding of Lipkus and others is that even well-educated participants perform poorly on tests of numeracy. For our purposes, one of their other results is more relevant, namely that a factor analysis that revealed a single factor was sufficient to characterize both general and expanded numeracy items. Lipkus and others concluded that existing measures of numeracy—that is, ones that are not necessarily posed in the context of the health domain—may be sufficient for assessing patients' ability to understand medical information. In short, their results imply that any distinction between general and health numeracy may not matter for practical purposes.

We revisit this conclusion by assessing more directly the potential for difference between health and general numeracy. In particular, we seek to determine whether numeracy is domain-specific by comparing participants' ability to carry out identical mathematical tasks with different contextual frames. In a nutshell, we ask whether people do better or worse when math problems are posed in the domain of health, compared with a financial domain or a purely mathematical one.

METHODS

Several members of the current study team (Fagerlin, Ubel, and Weir) led a team designing a data collection instrument on health numeracy that was included in the 2002 wave of the Health and Retirement Study (HRS).[21] The HRS is an ongoing, longitudinal, biennial study of 22,000 individuals ages 51 and older that was begun in 1992, with new sample cohorts enrolled every 6 years. Each survey wave includes 2 components: 1) a core set of questions asked of all participants and 2) supplemental questionnaires known as ''modules'' that are administered to random subsamples of approximately 1000 respondents.

Participants

The 2002 HRS sample represents the community-dwelling US population over age 54. Blacks, Hispanics, and residents of Florida are oversampled by design; the use of analysis weights that address unequal sampling probabilities as well as response rates that vary by racial and geographic subgroups yields nationally representative estimates.[22,23] Although the possibility of nonrandom attrition from the sample is a concern for any longitudinal study, several careful studies have documented that attrition bias in the HRS is not significant.[24–26] Table 1 supports this view by presenting evidence that the demographic characteristics—age, sex, marital status, race/ethnicity, and self-reported health status—of HRS core respondents closely match those of a similarly defined sample from the March 2002 Current Population Survey.

The 2002 HRS numeracy module was administered to a subsample of 1051 respondents who were randomly drawn from HRS core respondents who were not living in a nursing home and responded to the core survey themselves, as opposed to having a proxy provide responses. Of these, 962 completed the numeracy module; these 962 respondents are the participants in our study. Table 1 presents evidence that the demographic characteristics of our participants closely match those of similarly defined samples from the core 2002 HRS and the March 2002 Current Population Survey. Our participants are

Table 1 How Representative Are Our Study Participants? Comparison of Characteristics for Individuals Ages 55 and Older in Different Samples

	HRS 2002		CPS 2002
	(1) Numeracy Module Respondents	(2) All Core Respondents	(3)
Age	67.9 (9.5)	68.0 (9.6)	67.3 (8.4)
Female	0.61 (0.49)	0.56 (0.50)	0.55 (0.50)
Married	0.60 (0.49)	0.64 (0.48)	0.62 (0.49)
Race/ethnicity			
White non-Hispanic	0.85 (0.36)	0.83 (0.38)	0.84 (0.36)
Black non-Hispanic	0.08 (0.27)	0.09 (0.29)	0.09 (0.29)
Other non-Hispanic	0.02 (0.13)	0.02 (0.12)	0.07 (0.25)
Hispanic	0.05 (0.22)	0.06 (0.25)	0.06 (0.25)
Fair or poor self-reported health	0.25 (0.43)	0.27 (0.44)	0.29 (0.45)
Did not complete high school	0.21 (0.41)	0.25 (0.43)	0.24 (0.43)
Core numeracy (0–4)	1.5 (1.3)	1.3 (1.3)	N/A
Memory score (0–20)	9.8 (3.6)	9.0 (4.6)	N/A
Serial Sevens score (0–5)	3.4 (1.5)	3.5 (1.7)	N/A
Sample *n* (unweighted)	962	16,963	37,118

Note: N/A = not available. Results are presented as mean (standard deviation). HRS = Health and Retirement Study; CPS = Current Population Survey. Estimates are weighted using sampling weights. In each study, the sample is restricted to respondents ages 55 and older. Column (1) contains results for our study participants.

therefore likely to be representative of the community-dwelling US population over age 54, with the caveat that because of exclusion from the module of those who relied on a proxy respondent in the core survey, our participants may be in slightly better cognitive health than a truly representative sample.

All statistical analyses were performed using Stata version 12.1 (Stata Corporation, College Station, TX).

MEASURES

Domain-Specific Numeracy Measures (Experimental Module)

Our primary outcome variable was the accuracy of participants' answers to 4 mathematical questions posed in the experimental module. Several members of the current study team (Fagerlin, Ubel, and Weir) led the team that designed this module.[21] In the module, each of the 4 questions could be asked in 1 of 3 different domains: a ''pure math'' version, a health scenario, and a financial scenario. Table 2 displays the question text for all versions of each question. Respondents were randomized so that they answered 2 items in 1 domain and 1 item in each of the other 2 domains. For example, a respondent might be asked the ''pure math'' versions of items 1 and 3, the health version of item 2, and the financial version of item 4. This design eliminates the possibility that domain effects are in fact the result of differences in the underlying mathematical ability of the respondents who were asked different types of questions, because all respondents are asked all types of questions. Moreover, while all participants were asked the 4 items in the same order, the order in which domains were assigned to items was randomized across participants. For example, one participant might have received 1.math/2.financial/3.health/4.financial while another would randomly have received 1.financial/2.health/3.math/4.financial. This eliminates the possibility that the order in which the domains were assigned to items might bias the results (as would be the case if, for example, the health domain had always been assigned to the last item, with the order of the items varied). Thus, domain is randomized across participants and items, minimizing the potential for bias. Item nonresponse for these questions is between 4% and 6% for items 1, 2, and 3 and is 18% for item 4. We treat item nonresponse on these items as an incorrect response. This is consistent with how Lipkus and others[10] treated item nonresponse and also makes sense given both the higher probability of item nonresponse among participants with lower levels of education and the increasing probability of item nonresponse as the difficulty of questions increases. Item nonresponse is slightly higher in the pure math domain (14%) than in the financial (11%) or health (12%) domains; the difference in the nonresponse rate between the pure math

Table 2 Numeracy Questions Asked in 2002 Health and Retirement Study

Core Numeracy Questions: Administered to All Respondents

1. If the chance of getting a disease is 10%, how many people out of 1000 would be expected to get the disease?
2. If 5 people all have the winning numbers in the lottery and the prize is $2 million, how much will each of them get?
3. Let's say you have $200 in a savings account. The account earns 10% interest per year. How much would you have in the account at the end of 2 years?

Experimental Module Numeracy Questions:
Respondents receive only 1 version (math, market, or medicine) of each question.

Item	Math Domain	Financial Domain	Health Domain
1.	What is 15% of 1000?	A store is offering a 15% off sale on all TVs. The most popular television is normally priced at $1000. How much money would a customer save on the television during this sale?	A pill cures 15% of people who have a disease. If 1000 people have the disease and they all take the pill, how many people will be cured?
2.	The number 10 is what percentage of 1000?	If a customer saved $10 off a $1000 chair, what percentage would the customer have saved off the original price?	If the chance of getting a disease is 10 in 1000, what percentage of people will get the disease?
3.	Which of the following percentages is the biggest: 1%, 10%, or 5%?	Which of the following percentages represents the biggest discount in a sale: 1%, 10%, or 5%?	Which of the following percentages represents the biggest risk of getting a disease: 1%, 10%, or 5%?
4.	Which of the following is the most likely to happen: something that happens 1 in 100 times, something that happens 1 in 1000 times, or something that happens 1 in 10 times?	Which of the following represents the biggest chance of winning a lottery: a 1 in 100 chance, a 1 in 1000 chance, or a 1 in 10 chance?	Which of the following represents the biggest risk of getting a disease: a 1 in 100 risk, a 1 in 1000 risk, or a 1 in 10 risk?

and health domains is not statistically significant. The results reported below are largely unaffected if instead of treating item nonresponse as incorrect we drop observations with missing data.

Core Numeracy Measures (Core Survey)

All respondents in the core survey were asked 3 basic math questions, which are displayed in Table 2. The first of these is adapted from Lipkus and others[10] and the other 2 were developed for use in the English Longitudinal Study on Ageing (ELSA).[21,27,28] The core numeracy items were scored by giving respondents 1 point for each correct answer on questions 1 and 2; for question 3, respondents were given 1 point if they said "240," which is not quite correct but was the most frequent answer (given by 40% of respondents) and 2 points if they gave the correct answer of 242, which was given by only 11% of the sample. Summing scores on the 3 questions yields a core numeracy score from 0 to 4. This scoring method follows the practice of the ELSA investigators who developed these measures.[27]

The 3 core numeracy questions have relatively high rates of item nonresponse: 8%, 12%, and 35% for questions 1, 2, and 3, respectively. As above, we treat individual item nonresponse for these questions as incorrect responses. Alpha for the internal consistency of the 3 core numeracy items is 0.58 in our sample, comparable to the scores of 0.57–0.63 that Lipkus and others[10] report for their general numeracy scale measured across 3 different samples.

Measurement of General Cognitive Abilities

We use general cognition measures based on 2 tests administered in the core survey. The first of these is a word recall test in which respondents are read a list of 10 common words (e.g., *hotel, sky, water*) and are then asked to recall as many of them as possible both immediately after the list is read and also several minutes later. The total number of words the respondent correctly recalls at both opportunities, from 0 to 20, is a measure of memory. Respondents are also asked to count backward from 100 by 7s (100, 93, 86, etc.) up to 5 times, and the number of correct subtractions represents another measure of cognitive ability. We construct a cognitive composite with mean 0 and variance 1 by standardizing both variables, averaging them, and standardizing the result.

Demographic Variables and Measures of Socioeconomic Position

The HRS collects information from all core respondents on age, gender, race, Hispanic ethnicity, self-reported health status, and educational attainment. We characterize respondents' race and ethnicity using 4 mutually exclusive categories: white non-Hispanic, black non-Hispanic, other non-Hispanic, and Hispanic (any race). We also code educational attainment categorically: less than high school, high school graduate, some college, and education greater than or equal to a college degree. We use self-reported health status to create a dichotomous indicator that is equal to 1 if the respondent reports fair or poor health and 0 otherwise.

Analysis Plan

We first test the hypothesis that the domain in which a numeracy question is presented affects the probability of correct response. Specifically, we begin by presenting the average fraction of correct responses in each domain—math, financial, and health—and testing whether the fraction correct in the financial domain or the health domain differs significantly from the fraction correct when the question is asked in terms of pure math. We calculate these differences overall (pooling all 4 items) and separately for each item.

Next, we perform multivariate analyses that allow us to estimate simultaneously the effects of domain, item, and core numeracy on the probability of correct response. We estimate a logistic model with the outcome variable coded as 1 for correct and 0 for incorrect. To account for the potential correlation in the error term at the individual level (since each respondent contributes 4 observations to our data), we estimate the model using a generalized estimating equation (GEE); more specifically, we use Stata's *xtgee… family(binomial) link(logit)* command. The multivariate analyses are weighted using the analysis weights described above. We use this approach to estimate 3 nested models with progressively larger sets of explanatory variables. The first multivariate model includes only item (representing 1 through 4, dummy coded), math/financial/health domain (dummy coded), and core numeracy. The coefficient on the health domain dummy allow us to test the hypothesis that the probability of correct response in the health domain is the same as in the pure math domain; the coefficient on the financial domain dummy tests a similar hypothesis about the probability of correct response in the financial domain versus the pure math domain.

The second model interacts the domain dummies fully with the item dummies. This allows us to test the hypothesis that the probability of correct response in the health (or financial) domain is the same as in the pure math domain separately for each item. That is, are domain effects specific to certain items, or are they evident for all 4 items? Finally, we estimate a third model that augments these predictors with individual characteristics: gender, composite cognitive score, age, education (dummy coded representing less than high school graduate [omitted], high school graduate, some college, and college graduate or more), race, ethnicity, and a dummy for fair or poor health. The inclusion of these individual characteristics should not affect the estimated domain effects from the previous model, because of the randomized nature of the study design, but the effects of individual characteristics on the probability of correct response are interesting in their own right. In presenting the results of our multivariate models, we report average marginal effects and their standard errors calculated using Stata's built-in ''margins'' command for variables that enter the model directly (i.e., without an interaction term). As discussed by Ai and Norton,[29] standard errors on variables included in interaction terms must be calculated manually. We do this following the procedure described by Karaca-Mandic and others,[30(p262–3)] which involves calculating the difference in predicted probabilities as the interacted binary variables are changed from 0 to 1 while the other variables in the model are held constant at their means.

RESULTS

Table 3 reports the average fraction correct by domain and item. Overall, participants answered correctly 61.2% of the time. They were significantly more likely to answer questions posed in terms of pure math (66.3% correct) or in the financial domain (62.7% correct) than in the health domain (53.9% correct; significantly different from the pure math domain with $P < 0.001$). When we look at results separately for each item, the pattern just described is

Table 3 Fraction of Correct Responses by Domain and Item

	Domain		
	Math	Financial	Health
Overall	0.663	0.627 ($P = 0.063$)	0.539 ($P = 0.000$)
Item 1	0.631	0.660 ($P = 0.446$)	0.520 ($P = 0.007$)
Item 2	0.408	0.468 ($P = 0.111$)	0.252 ($P = 0.000$)
Item 3	0.896	0.842 ($P = 0.074$)	0.673 ($P = 0.000$)
Item 4	0.694	0.548 ($P = 0.000$)	0.724 ($P = 0.447$)

Note: Unweighted sample size is 3848 (962 respondents each asked 4 items). Means are weighted using analysis weights. The *P* value reported in each cell is associated with testing whether the fraction correct differs from the corresponding fraction for the pure math domain.

evident for items 1, 2, and 3. For item 4, however, respondents were not significantly less (or more) likely to answer correctly in the health domain compared with the pure math domain; the financial domain, in contrast, yielded significantly fewer correct responses to item 4.

Table 4 contains the multivariate logistic model results that allow us to estimate simultaneously the effects of domain, question item, and respondent characteristics on the probability of a correct response. The first column of Table 4 contains results from the most parsimonious model in which there are no interaction terms and no individual characteristics beyond core numeracy. This model suggests that on average, respondents are significantly less likely to respond correctly when questions are posed in the health domain, with a marginal effect of −0.161 points on the probability of correct response. The effect of the financial domain is not significant. The model also shows significant effects of item—not surprisingly, since some questions are harder than others—and also a significant effect of core numeracy. An additional point on the core numeracy scale leads to a significant increase of 0.126 in the probability of correct response, similar in magnitude to the effect of having the question posed in the pure math domain rather than the health domain.

The next column presents models that include interaction terms between domain and item. A chi-squared test confirms that these additional covariates significantly improve the fit of the model, with $P < 0.001$. Similar to the results presented in Table 3, we see a fairly consistent and significant negative effect of the health domain for items 1 through 3, ranging in magnitude from −0.164 to −0.276. As in Table 3, item 4 shows no effect for the health domain. The results for the financial domain are inconsistent. Question 2 is significantly easier for respondents when posed in the financial domain than the pure math domain, with an increase of 0. 087 in the probability of correct response, but the opposite is true for question 4, with a probability of correct response 0.144 *lower* in the financial domain than in the pure math domain.

Column 3 of Table 4 augments the model with individual characteristics; again, a chi-squared test confirms that these additional covariates significantly improve the fit of the model, with $P < 0.001$. As expected, given the randomized nature of the study, these additional covariates have little effect on the estimated domain effects or the interactions between domain and item. Adding these covariates does reduce the effect of core numeracy—a result likely explained by the fact that the vector of additional variables includes gender and education, both of which are significant predictors of numeracy—although the effect of core numeracy remains significant. The composite cognitive score also predicts a higher probability of correct response, while each year of age reduces the probability of correct response by six-tenths of a percentage point. Blacks and Hispanics are less likely to respond correctly, while being in fair or poor health has no significant effect on the probability of correct response.

DISCUSSION

The results of the current study indicate that domain matters. In particular, individuals do worse on quantitative tasks posed in the health domain than in terms of pure math or a financial domain. This pattern was evident for 3 of the 4 items we administered. This finding raises the possibility that health numeracy is a different construct from general numeracy and that it might predict behaviors—such as choices about medical decisions—differently than do other measures. Our current study does not attempt to test this possibility, but our findings suggest that future research on this topic is warranted. A potential explanation for the pattern of results that we observe on items 1 through 3 is that the value a person places on the outcome—even an outcome in a math problem—can influence his or her ability to give the correct response. This might also explain why item 4 shows a different pattern from the other 3 items; in item 4, the outcome in the financial scenario—winning the lottery—is significantly more positive, and unusual, than the outcomes described in the other domains, which are either neutral (in the pure math

Table 4 Multivariate Logistic Models: Marginal Effects (Dependent Variable = 1 If Correct Response)

	(1)	(2)	(3)
Main effects of domain			
Financial domain	–0.032 (0.020) P = 0.121	—	—
Health domain	–0.161 (0.023) P = 0.000	—	—
Main effects of item			
Item = 2	–0.253 (0.022) P = 0.000	–0.299 (0.034) P = 0.000	–0.336 (0.037) P = 0.000
Item = 3	0.209 (0.020) P = 0.000	0.207 (0.030) P = 0.000	0.233 (0.048) P = 0.000
Item = 4	0.035 (0.022) P = 0.112	–0.006 (0.035) P = 0.857	–0.006 (0.039) P = 0.874
Core numeracy	0.126 (0.011) P = 0.000	0.128 (0.011) P = 0.000	0.069 (0.012) P = 0.000
Domain effects, fully interacted with item			
Financial domain × item 1	—	–0.006 (0.043) P = 0.888	–0.007 (0.047) P = 0.872
Financial domain × item 2	—	0.087 (0.045) P = 0.050	0.097 (0.046) P = 0.037
Financial domain × item 3	—	–0.095 (0.045) P = 0.037	0.121 (0.057) P = 0.033
Financial domain × item 4	—	–0.144 (0.045) P = 0.001	–0.147 (0.052) P = 0.004
Health domain × item 1	—	–0.164 (0.044) P = 0.000	–0.181 (0.047) P = 0.000
Health domain × item 2	—	–0.173 (0.045) P = 0.000	–0.160 (0.081) P = 0.045
Health domain × item 3	—	–0.276 (0.052) P = 0.000	–0.327 (0.082) P = 0.000
Health domain × item 4	—	0.014 (0.050) P = 0.784	0.025 (0.055) P = 0.649
Additional covariates			
Female	—	—	–0.136 (0.025) P = 0.000
Composite cognitive score	—	—	0.083 (0.013) P = 0.000
Age	—	—	–0.006 (0.001) P = 0.000
Education = High school	—	—	0.089 (0.030) P = 0.002
Education = Some college	—	—	0.108 (0.035) P = 0.001
Education = College or more	—	—	0.261 (0.040) P = 0.000
Race = Black non-Hispanic	—	—	–0.166 (0.035) P = 0.000
Race = Other non-Hispanic	—	—	0.091 (0.108) P = 0.320
Hispanic	—	—	–0.117 (0.051) P = 0.026
Health is fair or poor	—	—	–0.049 (0.027) P = 0.066
Wald χ^2	350.50	360.24	575.57
Sample *n* (individuals)	962	962	962
Sample *n* (observations)	3848	3848	3848

Note: Means are weighted using analysis weights. Results are presented as marginal effect (standard error) and *P* value associated with H_0: marginal effect = 0.

scenario) or negative (in the health scenario, where the outcome is having a disease or taking a pill that is not very likely to cure the disease) relative to life circumstances. Perhaps even more important, whether or not health numeracy is a distinct construct, our study shows that many people struggle with mathematical tasks even more when confronting those tasks in a health domain than in a pure math domain. This means that in terms of individuals' ability to make informed decision about medical and health risks, the situation may be even worse than we thought based on most US adults' already poor performance at basic math tasks. The current policy emphasis on patient-centered care—as desirable as it may be for other reasons[31]—may have the unintended consequence of disadvantaging individuals with low numeracy. Our results illustrate the importance of figuring out better ways to present numbers to patients and the potential pitfalls of relying on studies that focus on explaining numbers in a general domain to inform the communication of numbers in a health domain.

Limitations

Our study has several limitations. First, our data are more than 10 years old. Although there is no reason to think that this biases the results, it would be desirable to replicate this study using more recent data. Second, the numeracy module that forms the basis of our study was administered to 1051 HRS respondents, but only 962 completed it (a 91.5% response rate for this component of the survey). Although these 962 respondents look very similar on observable dimensions to the full HRS sample, as shown in Table 1, we cannot rule out the possibility that our results are subject to nonresponse bias on other, unobservable dimensions. Third, the internal consistency of our measure of core numeracy, a key explanatory variable, is relatively low (alpha of 0.58). Finally, we administered only 4 items in the experimental module, and it is possible that the apparent effect of domain is specific to the mathematical tasks that these questions require. Moreover, the mathematical content of each item is not identical across the different domains, and this may have confounded the results. For example, in items 1 through 3, the financial domain version of the question involves calculating a percentage discount—a common shopping task—while the medical version requires the subject to calculate a risk or probability. This potentially confounds the conclusion that the results represent true domain effects, except in so far as the health domain inherently demands the use of probability or risk. A high priority for future work will be to expand our approach using more questions and involving a broader range of mathematical tasks.

REFERENCES

1. Nelson W, Reyna V, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. Ann Behav Med. 2008;35(3):261–74.

2. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull. 2009;135(6):943–73.

3. Hamm R, Bard D, Scheid D. Influence of numeracy upon patient's prostate cancer screening outcome probability judgements. Annual Meeting of the Society for Judgement and Decision Making. November 2003; Vancouver, British Columbia, Canada.

4. Peters E, Vastfjall D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychol Sci. 2006;17(5):407–13.

5. Ciampa PJ, Osborn CY, Peterson NB, Rothman RL. Patient numeracy, perceptions of provider communication, and colorectal cancer screening utilization. J Health Comm. 2010;15 Suppl 3: 157–68.

6. Cavanaugh K, Huizinga MM, Wallston KA, et al. Association of numeracy and diabetes control. Ann Intern Med. 2008;148(10): 737–46.

7. Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC. Literacy and numeracy skills and anticoagulation control. Am J Med Sci. 2004;328(2):88–93.

8. 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. Am J Public Health. 1997;87(6):1027–30.

9. Kutner M, Greenberg E, Baer J. A First Look at the Literacy of America's Adults in the 21st Century. Washington, DC: National Center for Education Statistics, US Department of Education, 2005. NCES 2006-470.

10. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21(1):37–44.

11. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127(11):966–72.

12. Fagerlin A, Zikmund-Fisher B, Ubel P, Smith D. Measuring numeracy when people hate math tests. Med Decis Making. 2003;23:560.

13. 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. Med Decis Making. 2007;27(5):663–71.

14. Huizinga MM, Elasy TA, Wallston KA, et al. Development and validation of the Diabetes Numeracy Test (DNT). BMC Health Serv Res. 2008;8:96.

15. Montori VM, Leung TW, Thompson CA, et al. Validation of a diabetes numeracy evaluation tool. Diabetes. 2004;53:A224–A5.

16. Apter AJ, Cheng J, Small D, et al. Asthma numeracy skill and health literacy. J Asthma. 2006;43(9):705–10.

17. Cosmides L, Tooby J. Cognitive adaptations for social exchange. In: Barkow JH, Cosmides L, Tooby J, eds. The Adapted Mind: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press; 1992. p 163–228.

18. Ancker J, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc. 2007;14: 713–21.

19. Montori V, Rothman R. Weakness in numbers: the challenge of numeracy in health care. J Gen Intern Med. 2005;20(11):1071–2.

20. Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke SE. A definition and operational framework for health numeracy. Am J Prev Med. 2005;29(4):375–6.

21. Ofstedal MB, Fisher GG, Herzog AR. Documentation of Cognitive Functioning Measures in the Health and Retirement Study. Ann Arbor (MI): Institute for Social Research, University of Michigan, 2005. DR-006.

22. Heeringa SGaJHC. Technical Description of the Health and Retirement Survey Sample Design. Ann Arbor (MI): 1995.

23. Ofstedal MB, Weir DR, Chen K-TJ, Wagner J. Updates to HRS Sample Weights. Ann Arbor: University of Michigan; 2011.

24. Michaud P-C, Kapteyn A, Smith JP, van Soest A. Temporary and permanent unit non-response in follow-up interviews of the Health and Retirement Study. Longit Life Course Stud. 2011;2(2):145–69.

25. Weir D, Faul J, Langa K. Proxy interviews and bias in the distribution of cognitive abilities due to non-response in longitudinal studies: a comparison of HRS and ELSA. Longit Life Course Stud. 2011;2(2):170–84.

26. Cheshire H, Ofstedal MB, Scholes S, Schroeder M. A comparison of response rates in the English Longitudinal Study of Ageing and the Health and Retirement Study. Longit Life Course Stud. 2011;2(2):127–44.

27. Huppert FA, Gardener E, McWilliams B. Cognitive function. In: Banks J, Breeze E, Lessof C, Nazroo J, eds. Retirement, Health and Relationships of the Older Population in England: The 2004 English Longitudinal Study of Ageing (Wave 2). London: The Institute for Fiscal Studies; 2006. p 217–42.

28. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127(11):966–72.

29. Ai CR, Norton EC. Interaction terms in logit and probit models. Econ Lett. 2003;80(1):123–9.

30. Karaca-Mandic P, Norton EC, Dowd B. Interaction terms in nonlinear models. Health Serv Res. 2012;47(1 Pt 1):255–74.

31. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood). 2010;29(8):1489–95.