Can Patients Interpret Health Information? An Assessment of the Medical Data Interpretation Test

Lisa M. Schwartz, MD, MS, Steven Woloshin, MD, MS, H. Gilbert Welch, MD, MPH

Objective**. To establish the reliability/validity of an 18-item test of patients' medical data interpretation skills. **Design**. Survey with retest after 2 weeks. **Subjects**. 178 people recruited from advertisements in local newspapers, an outpatient clinic, and a hospital open house. **Results**. The percentage of correct answers to individual items ranged from 20% to 87%, and medical data interpretation test scores (on a 0–100 scale) were normally distributed (median 61.1, mean 61.0, range 6–94).* Reliability *was good (test-retest correlation = 0.67, Cronbach's alpha = 0.71).* Construct validity *was supported in several ways. Higher scores were found among people with highest versus lowest numeracy (71 v. 36,* P *< 0.001), highest quantitative literacy (65 v. 28,* P *< 0.001), and highest education (69 v. 42,* P *= 0.004). Scores for 15 physician experts also completing the survey were significantly higher than participants with other postgraduate degrees (mean score 89 v. 69,* P *< 0.001). **Conclusion**. The medical data interpretation test is a reliable and valid measure of the ability to interpret medical statistics. **Key words:** numeracy; decision making; patient education. **(Med Decis Making 2005; 25:290–300)

Three forces are combining to promote informed patient decision making. First, there is growing professional consensus that patients need information to meaningfully participate in their own care, a position articulated by organizations such as the Institute of Medicine, Department of Veterans Affairs, and the US Preventive Services Task Force.[1–3] Second, there is considerable evidence that the public wants to be informed.[4–8] Finally, there are now countless mechanisms to provide the requisite information including public (e.g., the National Library of Medicine's MedlinePlus[9]) and private-sector Web-based resources; direct-to-consumer advertisements for drugs, tests, and surgeries; and patient decision aids.

Nonetheless, there are reasons to question a basic premise of informed decision making: namely, that the public is really able to understand the data presented.[10,11] To learn how well people understand medical information, investigators have begun to develop methods to define, operationalize, and measure the relevant skills. Some efforts have primarily addressed the role of functional health literacy—essentially, patients' ability to read the kind of information they see in typical health care settings such as on pill bottles and appointment slips.[12] Others have focused more on simple quantitative skills, sometimes called numeracy, essentially measuring the ability to perform simple mathematical operations (e.g., converting between percentages and proportions).[11,13] Recently, there has been a recognition that existing measures are too narrow and that better measures capturing higher level skills are needed. For example, Baker and others include the following item in the brief test of functional health literacy, testing the ability to use information: "Normal blood sugar is 60–150. Your blood sugar today is 160. If this was your score, would your blood sugar be normal today?"[12]

We know of no measure, however, for determining whether patients can use numbers to make comparisons, a fundamental requirement for informed decision making. We therefore developed a new measure that goes beyond testing simple literacy and numeracy and

The 1st 2 authors contributed equally to the creation of this manuscript—the order of their names is arbitrary.

Received 21 June 2004 from the VA Outcomes Group, White River Junction, VT (LMS, SW, HGW), and the Center for the Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH (LMS, SW, HGW). Revision accepted for publication 3 November 2004.

Address correspondence to Lisa M. Schwartz, MD, MS, VA Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT 05009; e-mail: Lisa.Schwartz@Dartmouth.edu.

DOI: 10.1177/0272989X05276860

examines the ability to compare risks and put risk estimates into context (i.e., to see how specific data fit into broader health concerns and to know what additional information is necessary to give a medical statistic meaning). Our measure covers the range of health information people routinely encounter in direct-to-consumer prescription drug advertisements, news media reports, and statements physicians might make to a patient. We believe the measure will help researchers assess the prevalence of such skills and help communicators design and evaluate more effective educational strategies to foster informed medical decision making. Herein we describe the development of the measure and how we established its validity and reliability.

METHODS

The goal of this study was to establish the basic attributes, reliability, and validity of a medical data interpretation test in a group of people with a wide range of quantitative abilities. To further establish validity, we also asked physician experts with strong critical reading skills to complete the survey.

Test Development

Our data interpretation test measures the ability to make sense of and compare medical statistics about disease risk and about risk reduction (see appendix). The formatted version of the test is available at http://129.170.61.41/research_tools.php. This involves having a knowledge base for comparisons, the ability to perform comparison tasks, and appreciating context (i.e., knowing what additional information is necessary to give a medical statistic meaning). We modeled our approach on the quantitative and document literacy segments of the National Adult Literacy Survey, which simulates real-world information people routinely encounter.[10] Our test uses examples in the form of direct-to-consumer prescription drug advertisements, news media reports, and statements a physician might make to a patient. We created hypothetical examples (instead of using existing materials) to ensure that participants had not encountered the information before. To avoid problems related to the variability of grading, all the questions are closed ended (i.e., multiple choice).

The medical data interpretation test consists of 18 items that are based on 20 questions. Fifteen of the 18 items map directly on individual questions. Three of the items are inferred from responses to 2 questions each: questions 7 and 20, 13 and 14, and 14 and 15, respectively (see appendix).

Table 1 Characteristics of Study Sample

	***N* = 178**
Mean age	55
(range)	(21, 85)
Sex (% women)	47
Household income (%)	
<$10,000	8
$10,000–$24,999	25
$25,000–$49,999	36
$50,000–$99,999	22
≥$100,000	9
Highest level of education (%)	
<High school graduate	7
High school degree	41
College degree	32
Postgraduate degree	20
National Adult Literacy Survey[10] (11 questions)	
0–5 correct	5%
6–8 correct	21%
9–11 correct	74%
Numeracy[11] (3 questions)	
0 correct	7%
1 correct	15%
2 correct	40%
3 correct	38%

Test Evaluation

To evaluate the data interpretation test, we recruited 178 people from advertisements in local newspapers (n = 98), the White River Junction VA outpatient clinic waiting area (n = 64), and a hospital open house at Dartmouth Hitchcock Medical Center (n = 16) (Table 1). Eighty-four (out of the 88 asked) completed the retest survey 2 weeks later. All respondents returned their survey by mail. All were paid $25.

We judged the performance of the data interpretation test using the following criteria.

Basic Attributes

Individual Items

For each of the 18 items, we calculated the proportion of responses left blank to measure question usability. To measure item difficulty, we measured the proportion of responses that were correct—where we treated missing as incorrect.

Aggregate Scores

To create scores, we counted up the number of correct answers and transformed the scores to a 0–100 scale. We examined the distribution of test scores overall and in various subgroups.

Reliability

Test-Retest Repeatability

We calculated Pearson correlation coefficients of the test and retest scores for the subset of 84 people who completed the retest survey 2 weeks later.

Internal Consistency

We measured internal consistency reliability with Cronbach's alpha, a measure of how responses to each question correlate with responses to all other questions. We sought items that measured the domain of data interpretation but included a wide range of skills (thus we hoped for a moderate Cronbach's alpha).

Validity

Content Validity

We developed our items based on reviews of the medical literature. We also sought feedback from individuals with expertise in statistics, education, cognitive psychology, and education. We revised the items based on this feedback. To formally assess content validity, we asked 20 Dartmouth Medical School faculty who teach evidence-based medicine (but were not involved in this study) to complete the data interpretation test and then formally rate its content validity using criteria derived from Feinstein's index of sensibility.[14] Specifically, they were asked to rate the clarity of the test items, how well the data interpretation test covers the important concepts in the domain of critical reading skills, and whether a person scoring poorly on the test would have very limited ability to interpret medical data. Fifteen physicians returned completed surveys.

Construct Validity

If our measure of data interpretation skills were valid, we expected that certain groups of people would perform better than others. Specifically, we tested the hypothesis that test scores were higher in people with higher quantitative literacy, higher numeracy, and higher educational attainment. Quantitative literacy was measured by a subset of 11 questions from the quantitative and document literacy sections of the National Adult Literacy Survey.[10] We then graded each question and counted the number of correct answers. Numeracy was measured by the total number of correct answers to a 3-item scale that we have used in prior work.[11] Educational attainment (defined as the highest level attained) was categorized as 4 levels: less than high school, high school graduate, college graduate, and postgraduate degree. In addition, we tested the hypothesis that the physician experts would score higher on the test than people with similar educational attainment (i.e., postgraduate degrees) but without the relevant expertise.

Analysis

We used *t* tests to compare group means and linear regression to test trends of test scores across quantitative literacy, numeracy, and educational attainment. We used STATA 8.0 (College Station, TX) for all analyses. All comparisons were 2-tailed and were considered statistically significant at $P < 0.05$.

RESULTS

Table 1 shows that the 178 participants represented a wide range of age, income, and formal education. In addition, participants also demonstrated a range of quantitative abilities as measured by our 3-item numeracy scale and a subset of quantitative and document literacy questions from the National Adult Literacy Survey.

BASIC TEST ATTRIBUTES

Individual Items

Tables 2 and 3 show the basic attributes of the individual items. Item nonresponse was low (ranging from 1% to 6%). Item difficulty was broad: the percentage of correct answers to individual items ranged from 20% to 87%. Question 3 was the most difficult; it asked "which additional piece of information would be the best evidence that Gritagrel (a hypothetical drug to reduce strokes) helped people?" Most people (71%) thought the correct answer was "fewer died from strokes in the Gritagrel group" rather than "fewer people died for any reason in the Gritagrel group." Item 10 was the easiest; this item was worded as follows: "Mrs. Jones is told she has a 28 in 1000 chance of dying from cancer . . . Mrs. Jones's doctor tells her that a new pill, CANCER will lower her chance of dying by 50%. Mrs. Jones decides to take the CANCER pill. Now, what is her chance of dying from cancer?" Eighty-seven per-

Table 2 Proportion of Correct and Missing Answers to the 18 Items on the Medical Data Interpretation Test for the 178 Participants

Question No.	Knowledge basis for comparisons	Answered Correctly	Answered Incorrectly: Completed	Answered Incorrectly: Left Blank
11	Know that a denominator is needed to calculate risk	75%	24%	1%
8	Know that denominators are needed to compare risks in 2 groups	45%	54%	1%
1	Know that the base rate is needed in addition to relative risk to determine the magnitude of benefit	63%	36%	1%
12	Know that a comparison group is needed to decide whether benefit exists	81%	18%	1%
3	Know that lowering all cause mortality provides better evidence of benefit than lowering a single cause of death	20%	79%	1%
	Comparison tasks			
6	Select "1 in 296" as larger risk than "1 in 407"	85%	14%	1%
	Inferred items[a]			
7&20	Rate the riskiness of a 9 in 1000 chance of death as the same as a 991 in 1000 chance of surviving	61%	37%	2%
13&14	Select a larger risk estimate for deaths from all causes than death from a specific disease	30%	69%	1%
14&15	Select a larger risk estimate for a 20-year than 10-year risk	39%	60%	1%
	Calculations related to comparisons			
10	Calculate risk in intervention group by applying relative risk reduction to a baseline risk	87%	11%	2%
9	Calculate 2 absolute risk reductions from relative risk reductions and baseline risks and select the larger	80%	19%	1%
17	Calculate relative risk reduction from 2 absolute risks	52%	46%	2%
18	Calculate absolute risk reduction from 2 absolute risks	77%	19%	4%
19	Calculate the number of events by applying absolute risk to number in group	72%	22%	6%
	Context for comparisons			
2	Know age and sex of individuals in the source data are needed	47%	51%	2%
4	Know that age of individuals in the source data is needed	60%	39%	1%
5	Know that risk of other diseases is needed for context	62%	35%	3%
16	Know that, for male smokers, the risk of lung cancer death is greater than prostate cancer death	60%	37%	3%

a. These items were based on a total of 5 separate questions.

cent were able to correctly select the answer "14 in 1000."

Aggregate Test Scores

Data interpretation test scores were created on a 0–100 scale. The observed scores were normally distributed with a mean score of 61 (standard deviation of 17)

Table 3 Psychometric Properties for the Medical Data Interpretation Test

Property	Measure Used	Scale Performance		
Basic attributes				
Individual items	Proportion not answering item (goal is to minimize item nonresponse)	1%-6% (mean 2%)		
	Proportion answering item correctly (goal is a broad range of item difficulty)	20% (item 3)-87% (item 10)		
Aggregate score	Mean score, range, and standard deviation	61 (s = 17, range = 6–94)		
Reliability				
Test-retest repeatability (whether the answers are the same over time)	Correlation of scores at test and retest 2 weeks later (goal is $r > 0.6$)	Pearson $r = 0.67$		
Internal consistency (degree to which items measure a single construct)	Cronbach's alpha (goal is 0.7–0.8; lower suggests lack of coherence as a scale; higher suggests items are redundant)	alpha = 0.71		
Validity				
Content validity (are relevant domains of "critical reading skills" covered?)	Initial questionnaire revised based on feedback from experts in education, statistics, and medicine.			
	Physician experts' responses to:	*Excellent/ Very Good*	*Good*	*Fair/Poor*
	How well would you say the critical reading test covers the important concepts inherent in "critical reading skills?"	60%	27%	13%
	How would you rate the clarity of the critical reading test?"	73%	20%	7%
		Strongly Agree		*Disagree/ Strongly Disagree*
	If a person got most of the critical reading test questions *wrong*, their ability to interpret medical research findings is probably very limited.	86%		13%
Construct validity (scale discriminates among groups with different attitudes or abilities)[a]	Comparison of mean scores with the following hypothesized relationships:			
	People with highest > lowest numeracy	71 v. 36	$P < 0.001$	
	People with highest > lowest literacy	65 v. 28	$P < 0.001$	
	People with highest > lowest education	69 v. 42	$P < 0.001$	
	Physician experts > other postgraduates	89 v. 69	$P < 0.001$	

a. Each comparison is between those with the highest versus lowest level of the characteristic (e.g., numeracy score of 3 v. score of 0).

and ranged from 6 (1 correct out of the 18 items) to 94 (17 out of 18 correct).

RELIABILITY

The medical data interpretation test demonstrated good repeatability (Table 3). The correlation of test and retest scores 2 weeks later was 0.67. The medical data interpretation test also has good internal consistency reliability (Cronbach's alpha = 0.71).

VALIDITY

The 15 physicians rated the content validity of the critical reading test highly (Table 3). Sixty percent thought the test did an excellent or very good job cover-

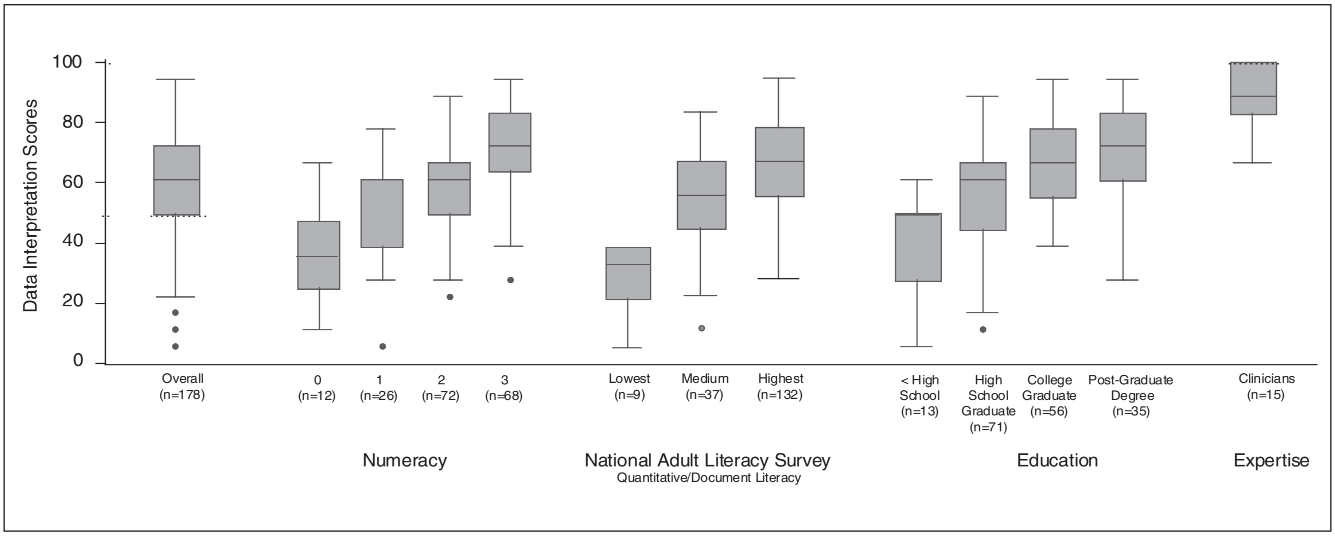

Figure 1 Distributions of scores of the medical data interpretation test for various subgroups. The lower end of each box is the 25th percentile, and the upper end is the 75 percentile (i.e., the box represents the interquartile range). The line in the middle of the box is the median. The lines emerging from each box extend to 1.5 times the interquartile range. Values beyond these lines are plotted individually. All tests of trend within subgroups are statistically significant (P < 0.001).

ing the important concepts in critical reading skills (27% rated the coverage "excellent"; 33% "very good"). Seventy-three percent rated the clarity of the test highly (20% rated the clarity "excellent"; 53% "very good"). Eighty-six percent agreed or strongly agreed that a person getting most of the questions wrong had very limited ability to interpret research findings (33% "strongly agreed"; 53% "agreed").

Figure 1 displays the distributions (medians with interquartile ranges) of data interpretation scores of various groups and supports the construct validity of our measure. People with high numeracy, quantitative literacy, and educational attainment all had higher critical reading scores than people with low levels of these characteristics (71 v. 36, 65 v. 28, and 69 v. 42, all $P < 0.001$). In addition, scores for physician experts were significantly higher than those of study subjects with postgraduate degrees (i.e., similar educational attainment but no relevant expertise): 89 v. 69, $P = 0.001$.

DISCUSSION

We created and validated a new measure to assess medical data interpretation skills. We designed the measure as a comprehensive test of skills needed to make sense of the kind of real-world health information people routinely encounter in direct-to-consumer prescription drug advertisements, news media reports, and statements physicians typically use in conversation with patients about medical risks. A group of physicians with expertise in evidence-based medicine reviewed the measure and supported its clarity, relevance, and content validity. We established that the measure had excellent reliability, usability, and construct validity in a study of men and women with a broad range of sociodemographic characteristics.

Our findings need to be interpreted in the light of several limitations. First, there may be disagreement on the content of the measure. Although we selected domains based on our own experience, a literature review, and discussions with experts in medicine, education, and psychology, others might have chosen differently. Similar concerns might be raised about individual items and their wording. Eight items may seem subjective because they include phrases like "What would you most like to know?" or "What information would best help you?" Strictly speaking, the answers are subjective in that people might want to know any of the response choices listed. However, we were careful in developing the questions to provide enough context to make the right answer reasonably clear (while still providing reasonable distracter alternatives). The responses of the 15 physician experts who completed the questions (to help establish construct validity) argue against subjectivity: on average, the experts answered 7 out of these 8 items correctly.

Second, as is always the case in the absence of an accepted gold standard, our validation approach is only as good as the constructs used to approximate the truth. We believe that the 4 hypotheses we prespecified (listed at the bottom of Table 3) represent sensible tests of construct validity. Third, the patients (and the physician experts) were a convenience sample. Nevertheless, although it is possible that other subjects might have responded differently, the main goal of our study was to examine the measures of performance in typical users. Finally, test usability was very good among study participants with less than a college degree: almost all of these respondents were able to complete every question. Nonetheless, our sample included few respondents with less than a high school education. Further study of test performance among those with very little formal education will be important.

Patients' desire for medical information is very strong,[4–8] and physicians, advocacy groups, pharmaceutical companies, and others are working hard—and effectively—to fulfill that desire. Ideally, if information were accurate and accessible, the result would be an increasingly informed public. In reality, people are exposed to messages of varying quality, with many messages serving to persuade rather than educate. We hope our new measure will help researchers understand how well the public is prepared to receive medical information. Specifically, we hope that it might be a useful baseline measure for studies assessing patient decision making. We also hope it will help inform and evaluate educational efforts to enhance the kind of data-interpretation skills needed to make informed medical decision making a reality.

ACKNOWLEDGMENTS

Drs. Schwartz and Woloshin were supported by Veterans Affairs Career Development Awards in Health Services Research and Development and by Robert Wood Johnson Generalist Faculty Scholar Awards. This study was supported by a grant from the National Cancer Institute (CA91052-01) and by a Research Enhancement Award from the Department of Veterans Affairs. The views expressed herein do not necessarily represent the views of the Department of Veterans Affairs or the United States government.

APPENDIX

Distribution of Answers for the 20 Questions and the 3 Inferred Items and Item-Score Correlations.
Bordered Cells Indicate the Correct Answer.

Question		(*N* = 178)	Item-Score Correlation
The next few questions are about the following advertisement for an imaginary drug called Gritagrel. (The following text appears in an ad graphic)			
"Gritagrel—50% reduction in strokes. Gritagrel is a new pill to prevent strokes. People taking Gritagrel had half as many strokes as people taking placebo (i.e. sugar pill). Like any medicine. Gritagrel can have side effects. The most common side effects are: headache (5%) and stomach ache (1%)."			
1. Which would best help you to determine how much a person could benefit from Gritagrel? (Circle one)			0.35
a. How often people experience side effects		11%	
b. The chance of stroke for people who do not take Gritagrel		[63%]	
c. How many people take Gritagrel		16%	
d. How recently Gritagrel was developed		9%	
	Left blank	1%	
2. Which would best help you to decide whether *you* will benefit from Gritagrel? (Circle one)			0.34
a. How many people were in the study		22%	
b. Age and sex of people in the study		[47%]	
c. Whether a doctor confirmed that people had strokes		12%	
d. Who paid for the study		17%	
	Left blank	2%	

(continued)

APPENDIX (continued)

Question		(*N* = 178)	Item-Score Correlation
3. Which additional piece of information would be the best evidence that Gritagrel helped people? (Circle one)			0.20
a. Gritagrel lowered antioxidant levels		1%	
b. Fewer people died for any reason in the Gritagrel group than in the placebo group		**20%**	
c. Many doctors prescribe it		7%	
d. Fewer people died from strokes in the Gritagrel group than in the placebo group		70%	
	Left blank	1%	
Your doctor says there is a 10% risk of dying of pneumonia			0.42
4. Which information best helps you understand how this risk *applies to you*?			
a. Most people who die from pneumonia are 75 years or older		**60%**	
b. More than 110,000 people get pneumonia each year		13%	
c. Pneumonia is one of the most common reasons for hospitalization		11%	
d. About 15,000 people die from pneumonia each year		16%	
	Left blank	1%	
5. To better understand how much of a threat pneumonia is to your health, which information is most helpful?			0.36
a. How much money is spent on pneumonia research		15%	
b. Whether pneumonia is more common in the US than Europe		18%	
c. Your chance of other important diseases		**62%**	
d. Celebrities who have had pneumonia		2%	
	Left blank	3%	
Mrs. Smith is told she has a 1 in 296 chance of dying from cancer and a 1 in 407 chance of dying from a stroke.			
6. Which is bigger, Mrs. Smith's chance of dying from a stroke or cancer?			0.47
a. Stroke		9%	
b. Cancer		**85%**	
c. Chances are the same		5%	
	Left blank	1%	
Mr. Roe needs surgery: 9 in 1000 people die from this surgery.	Scored with question 20 (see inferred items for distributions)		
7. How would you describe the surgery?			
a. Very risky			
b. Risky			
c. Slightly risky			
d. Not risky			
	Left blank	1%	
A new study finds that there were 30 deaths among people who eat broccoli regularly compared to 100 deaths among people who don't eat broccoli at all.			
8. According to this study, which statement best describes how eating broccoli relates to death?			0.23
a. Lowers the risk of death		49%	
b. Doesn't change the risk of death		2%	
c. Raises the risk of death		4%	
d. Can't tell from this information		**45%**	
	Left blank	1%	

(continued)

APPENDIX (continued)

Question		(*N* = 178)	Item-Score Correlation
Mrs. Jones is told she has a 28 in 1,000 chance of dying from cancer and a 59 in 1,000 chance of dying from a stroke. Mrs. Jones's doctor now tells her that a new pill, STROKE, will lower her chance of dying from stroke by 50%. Another pill, CANCER will lower her chance of dying from cancer by 50%.			
9. She can only take 1 pill. Assuming the 2 pills are equally safe and cost the same, which do you suggest she take?			0.51
a. STROKE pill		[80%]	
b. CANCER pill		18%	
	Left blank	1%	
10. Mrs. Jones decides to take the CANCER pill. Now, what is her chance of dying from cancer?			0.57
a. 0 in 1,000		3%	
b. 7 in 1,000		6%	
c. 14 in 1,000		[87%]	
d. 21 in 1,000		2%	
	Left blank	2%	
About 51,000 Americans will be diagnosed with melanoma (the most serious skin cancer) this year.			
11. What is your best guess about an American's chance of being diagnosed with melanoma in the next year?			0.47
a. 51,000		7%	
b. 51,000 divided by the number of Americans		[75%]	
c. Don't know how to figure this out		17%	
	Left blank	1%	

Promising new drug hailed*

Washington, DC—Researchers announced the results of a long awaited study of the drug Argentex, a drug designed to prevent prostate cancer. In the study, 1000 men age 45 to 75 were randomly assigned to take either Argentex or a sugar pill called a placebo. The men were followed for 4 years. Only 3 of the men taking Argentex developed prostate cancer. Lead scientist Bernard Womba described the findings as "extremely promising" and predicted the drug would be in wide use shortly.

*Not a real drug

Question		(*N* = 178)	Item-Score Correlation
12. Which question would you *most* like to ask Dr. Womba?			0.58
a. Who paid for the study?		8%	
b. Has Argentex been shown to work in animals?		3%	
c. What was the average age of the men in the study?		7%	
d. How many men taking the sugar pill developed prostate cancer?		[81%]	
	Left blank	1%	
Mr. Newman is a healthy 40-year-old man who does not smoke.			
13. What is your best guess about his chance of dying from a heart attack in the next 10 years?	Scored with question 14 (see inferred items for distributions)		
a. 1 in 1,000			
b. 5 in 1,000			
c. 30 in 1,000			
d. 80 in 1,000			
e. 250 in 1,000			
	Left blank	0%	

(continued)

APPENDIX (continued)

Question		(*N* = 178)	Item-Score Correlation
14. What is your best guess about his chance of dying for any reason in the next 10 years?	Scored with questions 13 & 15 (see inferred items for distributions)		
a. 1 in 1,000			
b. 5 in 1,000			
c. 30 in 1,000			
d. 80 in 1,000			
e. 250 in 1,000			
	Left blank	1%	
15. What is your best guess about his chance of dying for any reason in the next 20 years?	Scored with question 14 (see inferred items for distributions)		
a. 1 in 1,000			
b. 5 in 1,000			
c. 30 in 1,000			
d. 80 in 1,000			
e. 250 in 1,000			
	Left blank	1%	
Mr. Doe is a 75-year-old who smokes.			
16. How does his chance of dying from prostate cancer in the next 10 years compare to his chance of dying from lung cancer?			0.36
a. 10 times more likely to die of lung cancer than prostate cancer		60%	
b. Equally likely to die of lung cancer as prostate cancer		30%	
c. 10 times more likely to die of lung cancer than prostate cancer		6%	
	Left blank	3%	
In a new study, people either took pill X or placebo (a sugar pill). 3% of people taking placebo died; 1% of people taking pill X died.			
17. Which statement is correct about how pill X changes the chance of death?			0.57
a. Lowers by 66%		52%	
b. Lowers by 33%		38%	
c. Raises by 33%		4%	
d. Raises by 66%		3%	
	Left blank	2%	
18. Which statement is correct about how pill X changes the chance of death?			0.54
a. 2 more deaths per 100 people		7%	
b. 1 more death per 100 people		3%	
c. 1 fewer death per 100 people		8%	
d. 2 fewer deaths per 100 people		77%	
	Left blank	4%	
19. Suppose that 500 people had taken pill X, and 500 people had taken placebo in the new study. Which of the following is correct?			0.47
Number of people who died — *who took placebo* / *who took pill X*			
a. 15 / 5		72%	
b. 1 / 3		3%	
c. 3 / 1		14%	
d. 5 / 15		5%	
	Left blank	6%	

(continued)

APPENDIX (continued)

Question		(N = 178)	Item-Score Correlation
Mr. Smythe needs surgery. 991 in 1000 people survive this surgery			
20. How would you describe the surgery?	Scored with question 7 (see inferred items for distributions)		
a. Very risky			
b. Risky			
c. Slightly risky			
d. Not risky			
	Left blank	2%	
Inferred items			
Questions 7 and 20: Comparing the riskiness rating of a 9 in 1000 chance of death compared to a 991 chance of surviving			0.31
a. Chance of death rated riskier		25%	
b. Same riskiness		[61%]	
c. Chance of surviving rated riskier		14%	
	Left blank	2%	
Questions 13 and 14: Comparing the risk estimate selected for death for all causes to that for a specific disease			0.24
a. All cause risk estimate greater than disease specific		[30%]	
b. All cause risk estimate equals disease specific		62%	
c. All cause risk estimate less than disease specific		8%	
	Left blank	1%	
Questions 14 and 15: Comparing the risk estimate selected for a 20 year to that for a 10 year risk			0.21
a. 20 year risk estimate greater than 10 year		[39%]	
b. 20 year risk estimate equals 10 year		56%	
c. 20 year risk estimate less than 10 year		4%	
	Left blank	1%	

REFERENCES

1. Committee on Quality of Health Care in America. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
2. Department of Veterans Affairs. VHA Notice 99-02. Shared decision making. June 15, 1999. Available from: http://www.va.gov/putl/direc/health/notice/n9902.pdf. Accessed April 8, 2004.
3. Sheridan S, Harris R, Woolf S. Shared decision making about screening and chemoprevention: a suggested approach from the US Preventive Services Task Force. Am J Prev Med. 2004;26(1):56–66.
4. Nease RF, Brooks WB. Patient desire for information and decision making in health care decisions: the autonomy preference index and the health opinion survey. J Gen Intern Med. 1995;10:593–600.
5. Deber R, Kraetschmer N, Irvine J. What role do patients wish to play in treatment decision making? Arch Intern Med. 1996;156:1414–20.
6. Jones R, Pearson J, McGregor S, et al. Cross sectional survey of patients' satisfaction with information about cancer. BMJ. 1999;319:1247–8.
7. Laine C, Davidoff F, Lewis C, et al. Important elements of outpatient care: a comparison of patients' and physicians' opinions. Ann Intern Med. 1996;125:640–5.
8. Degner L, Kristjanson L, Bowman D, et al. Information needs and decisional preferences in women with breast cancer. JAMA. 1997;277:1485–92.
9. MedlinePlus. Trusted Health Information for You. A Service of the National Library of Medicine and the National Institutes of Health. Available from: http://www.nlm.nih.gov/medlineplus/
10. Kirsch IS, Jungeblut A, Jenkins L, Kolstad A. Adult Literacy in America. Washington, DC: National Center for Education Statistics; 1993.
11. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–72.
12. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Counsel. 1999;38:33–42.
13. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21:37–44.
14. Feinstein A. Clinimetrics. New Haven, CT: Yale University Press; 1987.