Final Progress Report

"Communicating Probabilities through Interactive Computer Graphics"

Principal Investigator: Rita Kukafka,[1,2]

Other team members: Jessica Ancker,[1] Elke Weber[3]

1. Columbia University College of Physicians and Surgeons, Department of Biomedical Informatics, 622 W. 168th Street, VC5, New York, NY 10034
2. Columbia University Mailman School of Public Health, Department of Sociomedical Sciences, New York, NY
3. Columbia University Department of Psychology, and Graduate School of Business Department of Management

Dates: July 1, 2006 – June 30, 2009
Federal Project Officer: Cindy Brach
Grants Management Specialist: Suzanne Holman
Funding: AHRQ 5R03HS016333-02

1. STRUCTURED ABSTRACT

Purpose: The ability to understand quantitative information about probabilities is critical for health literacy. Unfortunately, many people cannot understand or manipulate probabilities when they are presented as numbers, and people with poor numeracy have difficulty drawing emotional conclusions about quantitative risk information. This two-stage project was conducted to develop novel interactive computer graphics to communicate risks and assess their impact on decision-making and risk perception.

Scope: Health consumers were recruited through an urban community health promotion center for five focus groups to assist in developing new visuals. For a questionnaire study, consumers (n=100) were recruited through an online laboratory, and patients and families (n=65) were recruited at clinic waiting rooms at an urban hospital.

Methods: Focus groups using scenario-based usability testing, followed by a questionnaire study posing hypothetical health risks and a choice of whether to take a preventive action (such as a vaccine) to reduce that risk.

Results: Qualitative development methods with focus groups were used to develop an interactive Flash game that showed a matrix of square buttons; clicking on any button revealed whether the stick figure underneath was affected by the health outcome. In the focus groups, participants using this interaction to learn about a risk expressed more emotional responses, both positive and negative, than when viewing any static graphic or numerical description of a risk. Responses included relief about small risks and concern about large risks. The quantitative questionnaire study confirmed a strong correlation between the number of clicks performed in the interaction and perceived vulnerability to a risk; more clicks before finding a stick figure affected by the disease was associated with significantly lower perceived vulnerability. Numeracy strongly affected risk estimates, risk feelings, and decisions, with lower numeracy correlated with higher perceived risks. Interaction with one of the interactive graphics affected risk perceptions and narrowed differences between high- and low- numeracy respondents. Participants who used the interactive graphic to learn about the risk were more willing to take protective action against the risk than were participants who viewed noninteractive graphics illustrating the risk. The interactive graphics were rated as very helpful in understanding the risk, especially by participants who performed poorly on a numeracy scale. Interactive graphics may represent a novel way of exploiting computer gaming techniques to explain quantitative risks in health contexts.

Key Words: risk communication, graphics, interactivity, numeracy, health literacy

2. PURPOSE

Health communication and informatics interventions frequently require clear communication about the probabilities of health outcomes. Both health-related decision making (such as the choice between cancer therapies) and health promotion (such as decisions about vaccines and other preventive measures) require patients to use quantitative information. Tools that demonstrate risks clearly could help people of different educational levels communicate meaningfully about risks. Animation and interactivity are powerful computer tools that could be applied to this problem. Carefully designed interactive graphics might be particularly effective for people with low numeracy, because they might exploit associative learning and experience rather than effortful mental computation. Interactive communication and exploration can help people build more sophisticated understandings of complex concepts and engage decision processes similar to those they use in the real world.

This project involved two steps. A formative qualitative study was used to develop novel risk graphics. The objectives were to explore consumer preferences for different interactive graphics, basic usability, and consumer interpretations of the graphics and to use the results to develop prototypes for interactive risk communication modules that might be appropriate for web-based health promotion.

The objective of the second step was to assess the impact of these graphics in a quantitative questionnaire study focusing on decision making and perceived risk. Numeracy and health literacy were among the independent variables.

3. SCOPE

Background and context: Communicating about risks is an important part of health promotion, decision support, informed consent, and other health communication activities. Perceived risk is a potential motivator of health behavior change in many health models [1-3], and a recent meta-analysis confirmed that perceived risk has a strong relationship with behavior [4]. Patients should be able to understand and compare risks when making decisions, such as choosing between treatments [5], agreeing to screening tests [6], understanding insurance alternatives or healthcare quality indicators [7], and granting informed consent for treatment or research [8].

Unfortunately, communicating accurately about risks is difficult. Numeracy skills, such as arithmetic and ability to manipulate probabilities, vary widely among the public [9-11]. Numeracy and educational level are associated with comfort with [12, 13] and understanding of [10, 14] probabilities, percentages, and rates.

Graphs can be useful in illustrating risks [5, 15, 16]. However, graphical literacy skills often affect the ability to use information in graphs. For example, among samples from the general public, many were unable to calculate risks or differences between risks from survival graphs, a type of graphic used commonly for communicating with health professionals [17, 18].

Matrices of squares, icons, or stick figures depicting the people affected by the hazard are generally considered more intuitive and less dependent on graphical literacy. Even in such matrix displays, format can strongly affect behavior and judgment. For example, graphs that display the number affected (numerator) without showing the total number at risk (denominator) are more effective in promoting risk-reducing behavior [19, 20]. Highlighting the figures affected by disease in a random arrangement produced larger estimates of risk magnitude than did arranging them sequentially [21]. Elting et al found that physicians were more likely to act upon an association between a treatment and a negative outcome when data were presented as a matrix of squares than when they were presented as bar charts [22]. Interpretations and judgments of credibility can also be affected. Schapira et al found that some women viewing a matrix display thought the matrix showed the actual number of people from whom the information was derived, leading to judgments that the information was unreliable, because it derived from a small sample [23]. Matrix graphs with randomly arranged figures were judged more “true” than those with sequentially arranged figures [21]. As others have noted [21], more systematic exploration of graphics is needed to delineate these format effects on comprehension, decision making, and risk perception.

Graphics are often used to illustrate passive risk communications---that is, material to be read by patients or health consumers. By contrast, in active information processing, patients can engage with the information by manipulating it, elaborating on it, or acting on it. According to the elaboration likelihood model, attitude changes deriving from such active information processing will be more stable and persistent compared with changes deriving from peripherally processed information [24]. Comprehension and other outcomes may also be also affected. Emmons and colleagues found that a computer tool that allowed participants to input their own risk factors was effective in correcting misperceptions about personal risk for colorectal cancer [25]. In this study, patients using an active engagement version that allowed them to see the effects of altering their own modifiable risk factors were more accurate in their risk perceptions than were patients using a passive version, although the effect was not statistically significant [25]. Natter and Berry invited participants to draw a graph to illustrate a numerical risk or to answer a reflective essay question about it. Such active processing improved understanding of and satisfaction with the information [26].

A different perspective on the usefulness of active engagement with risk information comes from a recent set of choice behavior studies [27, 28]. In these studies, experimenters assembled two decks of cards, each displaying the amount of money to be won by drawing that card. Participants could sample as many cards from the two decks as desired, then choose a deck to draw a card for a real monetary payoff. Decisions in this "learning from experience" paradigm were quite different from those made when risks were described in numbers or graphs. For example, as described in prospect theory [29], people usually overweight the probability of rare events (for example, avoiding gambles that carry a small chance of losing a large sum, even if on average, such a gamble would be a good bet). However, they underweight the probability of rare events when learning from experience [27, 28]. This effect has been suggested as a source of communication barriers between doctors, who may learn about medical events through personal experience, and patients, who learn about them through description [27].

Participants and settings

For the qualitative study, health consumers were recruited through fliers at public clinics, libraries, recreation centers, and other organizations in the Harlem neighborhood of New York City. Focus groups were conducted at the Harlem Health Promotion Center. All participants provided written informed consent and were reimbursed with movie ticket vouchers and subway passes.

For the questionnaire study, two groups of participants were recruited. For the online sample, we collaborated with the Columbia Center for the Decision Science's Virtual Lab, a pool of >30,000 volunteers around the country who have registered to participate in short, online, decision-making studies. Recruitment took place from this pool from July 1 through July 18, 2008. In total, 101 people responded with complete data, but one was dropped from the analysis for completing the questions in an obviously rushed (<6 minutes total) and nonsensical manner (score, 5 of 36 on the TOHFLA and 1 of 8 on numeracy). Online participants were reimbursed through PayPal.

For a clinic-based sample, 65 participants were recruited from clinics at NewYork-Presbyterian Hospital, an urban teaching hospital in northern Manhattan. We recruited from waiting areas (the pediatric emergency department, the pediatric asthma room, and the dental teaching clinic) from July 30 to October 22, 2008. The researcher brought a laptop on a wheeled stand to the waiting areas and distributed flyers to all people in the waiting area. Volunteers completed the questionnaire on the laptop. All were reimbursed with movie ticket vouchers.

4. METHODS

Qualitative methods

We sought to combine the advantages of focus group research with those of a software development method called scenario-based usability engineering.[30 31, 32] In scenario-based usability engineering, focus group participants meet with developers to discuss realistic stories ("scenarios") describing individuals encountering a problem and using a computer system to resolve it. The process maximizes users' contributions but is not a controlled usability experiment. Small sample sizes are justified by Nielsen's sample size formula for usability testing: $P = 1 - (1 - p)^n$, in which P is the proportion of problems found, p is the probability an individual will find a problem, and n is the sample size. [32]

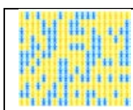
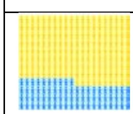
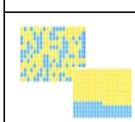
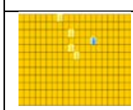
Under the assumption that any individual has a roughly 30% chance of finding a particular usability problem (an assumption based on Nielsen's empirical findings with relatively simple systems), a sample as small as 3 can be expected to identify 67% of usability problems, and a sample of 5 can be expected to catch 84% of problems.[32] All discussions were professionally transcribed and then coded for qualitative themes.[33] Two coders developed a codebook, annotated the transcripts individually, and reached consensus on the line-by-line coding of each transcript. Transcripts were annotated using ATLAS.ti™ software (ATLAS.ti, Berlin, DRG).

Focus groups were invited to discuss a "personalized risk scenario," modeled after stimuli used in recent studies of graphics.[21, 23, 34] In this scenario, a character named "Michelle" goes to a doctor, who examines her, states her risks of developing heart disease and breast cancer (e.g., "Your lifetime risk is 46%"), and gives her advice about a heart-healthy lifestyle. Focus group members were asked to imagine that "Michelle" has asked them for help understanding the risk. For three of the focus groups, we also created an interactive cardiac risk calculator based on the National Cholesterol Education Program's guidelines[35] so that participants could input personal characteristics (age, sex, cholesterol level, etc.) and receive a numeric risk similar to the one in the Michelle story as well as explore the effect of changing variables, such as cholesterol level. Participants were also invited to discuss the utility of illustrating the risk with three static graphics: a simple bar chart, a matrix of stick figures with affected people scattered randomly throughout the group, and a matrix of figures with those affected arranged sequentially. Participants then used laptops to explore the interactive risk communication module under development.

Quantitative methods

An additional 165 participants were recruited from the two settings (online lab, and urban hospital waiting rooms) to complete a questionnaire. All participants read two short stories about being faced with a health decision, one in which the probability of disease was 29% and one in which the probability was 6%. Each was offered a preventive action. However, the preventive action carried a risk of side effects that was about 1/3 the probability of disease without the preventive action (9% for the 29% story, and 2% for the 6% story). This design was meant to reduce the possibility of ceiling effects by ensuring that not all readers would choose the preventive action.

Each story was illustrated with a graphic based on a grid of figures, with the figures affected by the health condition colored blue and the rest colored yellow. The graphics were designed after the completion of the qualitative study (see Results section below). The questionnaire software assigned participants to four groups.

	In Group 1, both stories were illustrated with *random* graphics; that is, the blue figures were scattered randomly throughout the grid.
	In Group 2, both stories were illustrated with *sequential* graphics; that is, the blue figures were sequentially arranged in the bottom rows of the grid.
	In Group 3, stories were illustrated with interactive *switch* graphics, which allowed participants to switch back and forth between sequential and random views of the same percentage. Participants were required to switch states at least twice before being able to answer the questions.
	In Group 4, stories were illustrated with interactive *search* graphics, in which participants had to click on squares to see the color of the figures underneath. Once a blue figure was revealed, all the other squares turned over to reveal all the figures. Participants were required to find a blue figure before being able to answer the questions.

Each story was then followed by 11 questions about perceived threat, perceived efficacy, and intention to get the preventive action. Half the respondents in each group received the questions in reverse order as an order control.

The major constructs assessed were:

- feeling of risk (perceived susceptibility);
- risk estimate (outcome expectation);
- decision/choice (behavioral intention).

The term "feeling of risk" is taken from [36]; the Weinstein language will be used in the question, because he has recently demonstrated a strong correlation between the feeling of risk as measured with his questions and subsequent behavior (flu vaccination). The "feeling of risk" construct is the same as to "perceived susceptibility" in the health belief model [1].

The risk estimate or outcome expectation is taken from risk graphics studies, which have focused on the relationship between graphically portrayed risks and viewers' quantitative estimates of those risks [15]. Also inspired by risk graphics studies is the construct of the perceived credibility and accuracy of the illustration [21].

Behavioral intention was measured as a proxy for actual behavior. Additional health belief model constructs were included as covariates (perceived severity, perceived barriers, and self efficacy) [1].

Health literacy was assessed with the Short Test of Functional Health Literacy in Adults (S-TOFHLA), a validated test of health literacy that requires approximately 7 minutes to complete [37, 38]. Numeracy will be assessed with a short scale that has been normed in other samples by Lipkus [10, 11]. The scale requires a few minutes to complete and supplements the S-TOFHLA questions by focusing on probability, proportions, and percentages.

The online instrument was pilot tested for usability with seven participants of varying computer familiarity. After adjustments to improve usability for the lower familiarity participants, it was assessed for test-retest reliability. Test-retest reliability was assessed by having an additional nine participants take the questionnaire twice over an interval of 2 to 3.5 weeks (median: 18 days, range: 15-25 days). All were assigned to Group 4 (the interactive "search" graphic). It was anticipated that item correlations for the "decisions" subscale, which included all the perceived risk questions as well as the decisions, would be somewhat low because of the variability induced by the variability in their interactions with the graphic. In the analysis, average item correlations for the subscales were very high (all rhos >0.91), and the "decisions" subscale had a rho = 0.61 ($p < .001$).

5. Results

Principal findings of qualitative study

Several major themes arose in the qualitative analysis.

■ *Emotional impact:* The interactive program was associated with more expressions of emotion by the participants than the other graphics and visuals (for example, 17 comments were coded as emotional responses for the interactive program, compared with four for the bar chart). Several participants expressed dismay when a stick figure icon "got the disease." One woman said she didn't want to "play" anymore. Conversely, a participant who was exploring a low risk said she was relieved because she clicked so many times without finding a figure with the disease. The participants rarely expressed any emotional reaction to the printed illustrations of stick figures or to the bar chart.

■ *Stick figures:* Most participants preferred the matrix of stick figures to the bar chart, because it was "clearer that you're talking about human beings and not statistics." However, other participants felt that the stick figure display was somewhat overwhelming and "a lot to look at" and that the bar charts were "straightforward." Most thought the random arrangement was more difficult to count than the sequential arrangement. However, positive comments about the random arrangement outnumbered negative ones (8 positive to 2 negative). Many described it as realistic (e.g., "the more realistic way to depict the chance"). One said that the sequential arrangement "can give a person a false reading," because it made it too easy for her to imagine that she would *not* be affected by the risk.

■ *Interaction:* The interactive process of clicking squares to see the icons underneath was interesting to some of the participants. One said, “It’s like a game because you’re playing around with it. That’s what I like about it. Because you learn, too.” However, some others were impatient with the interaction, either because they said they already understood the percentage or because the module emphasized the negative (getting the disease) too strongly. Most said they enjoyed the interactive aspect of being able to input personal information and get tailored output.

Participant suggestions and observations of participant behavior were used to update and develop the software prototype throughout the project. Examples of changes to the prototype include:

■ *Stick figure arrangement:* All groups saw both randomly arranged and sequentially arranged stick figure illustrations. As described, participants generally agreed that it was easier to judge the proportion in the sequential arrangement of stick figures but described the random arrangement as more “realistic.” The final version displayed the sequential arrangement first so that viewers could easily see the proportion. When viewers advanced to the next screen in the module, they saw the random arrangement of the same quantity.

■ *Interaction:* The prototype for the first three groups showed a grid of squares with one of them colored blue. The user was invited to imagine that he or she was standing underneath it; clicking that square revealed all the stick figures, including the one representing “you.” The user could then reset the grid of squares and try again. As mentioned, this sometimes elicited a strong emotional response. We were concerned that, if the image was too threatening, people might refuse to use the program or reject its message [3]. To attenuate the impact somewhat, we changed the wording after the third focus group so that the illustration was described as showing someone among “a group of people” rather than “you.” This appeared to reduce the negative emotional responses in later groups without eliminating them. After reviewing the final focus group transcript, we also developed a modification that would allow users to click on multiple squares rather than on only a single square. The goal was to maximize interaction with the graphic while still satisfying participants’ desire to reach the “action steps” advice quickly, but this version was not assessed.

The continuous S-TOFHLA score was weakly but significantly correlated with the continuous numeracy score ($r = 0.28$, $p < .001$). However, low numeracy remained prevalent in the high-literacy group (28% of those with adequate health literacy had low numeracy).

Intention

Overall, the preventive medication was chosen by 62% of respondents in Story 1 and 49% of respondents in Story 2. Group 4 respondents were more likely than those in the other groups to opt for the preventive action in Story 1 (77% vs. less than 60% for other groups; $p = 0.11$) but not in Story 2.

Blacks and Hispanics were much more likely to opt for the preventive action than were Whites or Asians (Story 1: 75%, 78%, 53%, 35%; $p = .02$; Story 2: 55%, 64%, 36%, 35%; $p = .058$). Also, the clinic respondents were more likely than the online sample to opt for the preventive action (71% vs. 56% for Story 1; $p=.056$; 60% vs. 41% for Story 2; $p=.02$). Racial differences partly mediated the clinic effect (i.e., adding clinic to a logistic regression equation) reduced the coefficient of the race variable and increased its p value.

For Story 2, there was a strong question order effect. In the forward order, questions about perceived risk appeared immediately after the interaction with the graphic and before the question about intention to take the preventive action. In the reverse order, questions about disease severity and efficacy appeared first before the intention question, with the perceived risk questions afterward. For Story 2, reverse-ordered groups were much more likely to opt for the preventive action (55% or more versus 43% or less; $p = .01$). There was no similar order effect for Story 1.

Self-reported history of flu was not associated with likelihood of choosing the vaccine in Story 1, and self-reported history of heart disease was not associated with likelihood of choosing the preventive medicine in Story 2. Self-reported history of drug side effects was associated with a slightly lower likelihood of choosing the vaccine in Story 1 (53% vs. 68%, $p = .07$) but not in Story 2. Health status and educational level were not associated with likelihood of choosing the preventive action.

In both stories, all answers about perceived risk, self efficacy, response efficacy, and response side effects were correlated with intention to opt for the preventive action, in the expected directions, as predicted by the extended parallel process model. The numeric risk estimate was also correlated with intention. However, perceived disease severity was not significantly correlated with intention (Story 1: rho = 0.10; p=.2; Story 2: rho = 0.12, p = .14).

Principal findings of questionnaire study

Table 4.1 - Characteristics of online and clinic study samples

Characteristic	On-line (n=100)	Clinic (n=65)	p	Total sample (n=165)
Mean age, yrs (range)	32.8 (19-61)	30.7 (18-72)	.90	32.0 (18-72)
Number (%) women	64 (64.0)	41 (63.1)	>.99	105 (63.6)
Educational level, n (%)				
no bachelor's degree[a]	19 (19.0)	28 (45.0)	<.001	47 (28.5)
some college	37 (37.0)	23 (35.4)		60 (36.4)
bachelor's or graduate degree	44 (44.0)	14 (21.5)		58 (35.2)
Self-identity, n (%)				
African-American	10 (10.0)	10 (15.4)	<.001	20 (12.1)
Asian	20 (20.0)	0		20 (12.1)
White	60 (60.0)	6 (9.2)		66 (40.0)
Hispanic	2 (2.0)	43 (66.2)		45 (27.3)
other	3 (3.0)	3 (4.5)		6 (3.6)
mixed race/ethnicity	5 (5.0)	3 (4.5)		8 (4.8)
Mean health status ± SD (1 = poor, 5 = excellent)	4.0 ± 0.7	3.7 ± 0.8	.002	3.9 ± 0.7
Computer questions, n (%)				
use every day	98 (98.0)	37 (56.9)	<.001	135 (81.8)
very comfortable using mouse	96 (96.0)	56 (86.2)	.04	152 (92.1)
have no email address	0	7 (10.8)	.001	7 (4.2)
Numeracy score out of 8, n (%)				
<=5	16 (16.0)	31 (50.8)[b]	<.001	47 (29.2)[b]
6 or 7	47 (47.0)	27 (44.3)		74 (46.0)
8	37 (37.0)	3 (4.9)		40 (24.8)
S-TOFHLA category, n (%)				
adequate health literacy	99 (99)	56 (98)[c]	>.99	155 (99)
marginal health literacy	1 (1)	1 (2)		2 (1)

Chi-squared tests used for categorical variables and t tests for continuous ones.
a. Includes less than high school, high school graduate, and technical school
b. 4 clinic respondents missing numeracy scores because of interruptions
c. 8 clinic respondents missing TOFHLA scores because of interruptions
d. Time minus interruptions for registration, triage, dental appointment, etc.

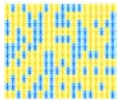
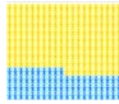
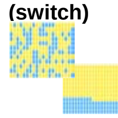
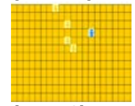
Interactions and attitudes toward graphics

The *search* graphic in Group 4 prompted a substantive amount of interaction. In Story 1 (involving the 29% risk), participants clicked a median of two times (range 1-16) before discovering a blue figure. In Story 2 (a 6% risk), participants clicked a median of 12 times (range 2-51).

In Group 3 (switch), respondents were required to switch twice between sequential and random views of the same percentage and then were free to interact more. Only 13 (33%) explored the graphic more, with a mean of 3.0 (sd = 1.8) clicks for the first story and 2.7 (sd = 1.2) for the second story.

In general, participants appeared to like the interactive graphics, particularly the searching graphics, although this version also earned the most "confusing" ratings (Table 5.2).

Table 5.2 – Attitudes toward the graphics (univariate analyses)

Strongly agreed that the graphic...	**Group 1 (random)**  **(n = 39)**	**Group 2 (sequential)**  **(n = 44)**	**Group 3 (switch)**  **(n = 39)**	**Group 4 (search)**  **(n = 43)**	**p**
Helped me understand the risks					
all participants	13/39 (33%)	18/44 (41%)	17/39 (44%)	**26/43 (60%)**	.08
low numeracy only	6/13 (46%)	6/12 (50%)	6/13 (46%)	**7/9 (78%)**	.46
Was confusing					
all participants	1/39 (3%)	0	1/39 (3%)	**4/43 (9%)**	.13
low numeracy only	0	0	0	0	
Graphic in Story 1					
Was realistic					
all participants	6/39 (16%)	14/44 (32%)	13/39 (33%)	**23/43 (53%)**	.004
low numeracy only	4/13 (31%)	4/12 (33%)	3/13 (23%)	**7/9 (78%)**	.07
Was an accurate way of showing the risks					
all participants	7/39 (18%)	15/44 (34%)	**19/39 (49%)**	**18/43 (42%)**	.03
low numeracy only	4/13 (33%)	4/12 (33%)	4/13 (31%)	**6/9 (67%)**	.33
Graphic in Story 2					
Was realistic					
all participants	13/39 (33%)	18/44 (41%)	13/39 (33%)	**21/43 (49%)**	.42
low numeracy only	4/13 (31%)	6/12 (50%)	4/13 (31%)	**6/9 (67%)**	.27
Was an accurate way of showing the risks					
all participants	12/39 (31%)	20/44 (46%)	18/39 (46%)	**24/43 (56%)**	.15
low numeracy only	5/13 (39%)	6/12 (50%)	5/13 (39%)	**5/9 (56%)**	.81

Numeracy and risk perception

Among respondents without a college degree, lower numeracy score was correlated with higher risk feelings and numeric risk estimates for both scenarios. Low education level (less than a bachelor's degree) was also associated with higher risk feelings and numeric risk estimates; the correlations between numeracy and risk perceptions were much weaker or nonexistent among those with bachelor's or advanced degrees.

Low-numeracy respondents were less likely than high-numeracy ones to choose the scenario-provided risk level for their numeric risk estimate; 67% of low-numeracy respondents gave a non-scenario risk estimate for both stories compared with 47% of high-numeracy ones (p=.02). Again, the effect was weaker among college-educated respondents.

Relationship between numeracy and attitudes toward graphics

People with poorer numeracy tended to consider graphics in general more helpful for understanding (i.e., numeracy score correlated with better helpfulness rating, rho=0.24, p = .002; 53% of low-numeracy respondents strongly agreed graphics were helpful compared with 42% of others). People with lower numeracy were not more likely to consider graphics confusing (23% vs. 19%, p=.56). As a validity check, ratings on the helpful and confusing questions were negatively correlated (rho= −0.37, p<.001). Low-numeracy respondents were more likely than high-numeracy ones to opt for the preventive action in Story 2 (60% vs. 44% for Story 2; p = .07).

Respondents with lower familiarity with computers (i.e., self-reported less frequent computer use and/or lack of an email address) did not give substantively different answers to the attitude questions, with one exception. Two of 7 (29%) considered the Group 4 graphics confusing compared with seven of 36 frequent computer users (19%; p = .03).

Effects of graphics on risk perception

Group assignment (i.e., graphics) did not affect median risk feelings, verbal risk estimates, or numeric risk estimates, but it did affect variability in numeric estimates. For Story 2, variance in numeric risk estimates was highest for the random graphic, lowest for the sequential graphic, and in the middle for the switch graphic and for the search graphic. The variances were significantly different (Leven's test, F=4.7, p=.004). The differences were attributable to more extreme outliers in the random group. In story 1, variances followed the same pattern but differences were smaller and not statistically significant.

Effect of the *search* graphic on perceived risk and intention to take preventive action

As noted above, respondents in Group 4 (*search* graphic) were more likely to take the preventive action than respondents in the other groups. The interaction itself had interesting effects.

For Story 2, with the 6% risk, the *search* graphics participants clicked from two to 51 times. There were fairly strong correlations between numbers of clicks and perceived risk, but even stronger with question order effects. However, click number and perceived risk did not strongly affect intention.

Half these participants received the questions about perceived risk immediately after their interaction with the graphic, before being asked about their intention to take the preventive action (forward order). With these respondents, number of clicks was correlated with perceived vulnerability (rho = −0.47; p=.04) and qualitative risk (rho = 0.57, p=.009), and there was also a small correlation with perceived susceptibility (rho = −0.22, p=.35). That is, the more clicks before finding a blue person, the lower the perceived risk. However, only six (30.0%) opted for the preventive action, and there was no correlation between intention and click number (rho = −0.02, p=.93).

The other half of Group 4 received the questions about disease severity and efficacy before making the decision, with the perceived risk questions afterward (reverse order). In this subgroup, neither perceived risk questions (vulnerability, susceptibility, and qualitative risk) nor intention were correlated with click number (all rhos < .13). However, a much larger proportion of this subgroup opted for the preventive action (75% of reverse-order respondents compared with 30% of forward-order ones, p=.004).

There appears to be some evidence that the *search* interaction may be more effective at reducing perceived risk than at increasing it. That is, in the subgroup that clicked many times (>= 25) before finding a blue person, none (0%) reported feeling susceptible to the disease. By contrast, in the subgroup that clicked only a few times before finding a blue person (<=10), the proportion who felt susceptible was virtually the same as the proportions in the other experimental groups (18% compared with 16%, 26%, and 26%; p =.67). Effects were similar but weaker with the vulnerability question; effects with the qualitative risk question were mixed.

In Story 1 (with the 29% risk), participants in the *search* group clicked a median of only two times before finding a blue figure, and there were no correlations between click number and the perceived risk questions (all rho ≤ 0.07).

Implications

Interactive graphics may represent a novel way of exploiting computer gaming techniques to explain quantitative risks in health contexts.

The qualitative work allowed us to pilot different types of game-like interactive risk graphics, reject types that were unlikely to be acceptable, and explore participants' interpretations of what they were seeing and experiencing. We designed the study to examine usability, user preferences, and user interpretations and to incorporate these findings into software modifications. The designs seemed to be highly usable. The game-like interaction of clicking on squares to look at the people underneath seemed to be more engaging and interesting. One unexpected and promising finding was that this interactive software appeared to produce a strong emotional response to the interactive graphics that was not found with any static graphic or number. This was especially intriguing in light of work by other researchers (e.g., [39]) that it is particularly difficult for people with poor numeracy to make emotional sense of numerical information.

The questionnaire study confirmed a strong correlation between the number of clicks performed in the interaction and perceived vulnerability to a risk. Participants who used the interactive graphic to learn about the risk were more willing to take protective action against the risk than participants who viewed noninteractive graphics illustrating the risk. The interactive graphics were rated as very helpful in understanding the risk, especially by participants who performed poorly on a numeracy scale. They were not significantly more likely to be considered confusing.

(a) In the searching interaction, the more the respondent clicked before finding a blue person, the lower the perceived personal risk. However, this effect was short term, disappearing after respondents were asked additional questions about different dimensions of the decision.

(b) In this short-term effect, a long sequence of clicks tended to lower perceived risk below the level induced by other graphics, but a short sequence of clicks (finding the blue person immediately) did not raise perceived risk above that level. This is particularly intriguing in light of findings by other researchers that it is especially difficult to reduce people's tendency to overreact to very rare risks (e.g., [40]). Interactivity of the type explored in this project may be more helpful, in a public health context, in calming fears about rare events (such as change of adverse effects from a beneficial vaccine) than in increasing attention to common threats (such as the threat of heart disease).

(c) Number of clicks in the interaction was not directly correlated with intention, but the effect was mediated by effect on perceived risk.

(d) The questionnaire study found that using question order to draw attention first to the severity of the disease and the effectiveness of the drug strongly increased intention to take the preventive option.

As expected, numeracy had strong effects in this study. Although numeracy was associated with lower educational level, our regression models suggest that the numeracy effects were not fully explained by education, age, computer literacy, or other factors. In the scenario study, low-numeracy respondents systematically gave numeric risk estimates that were higher than the value described in the scenario as well as higher than the estimates provided by other respondents. Low-numeracy respondents also reported higher risk feelings and higher verbal risk estimates than other respondents. This effect was in part because low-numeracy respondents were more likely to choose 50% when describing their own risk. This may be a rhetorical measure to express uncertainty; however, this phenomenon did not fully account for the overestimation, because average estimates were above the scenario risk even among participants who did not choose 50%. Another reason low-numeracy respondents' estimates were higher than high-numeracy respondents' was that, in the low-risk scenario, high-numeracy respondents tended to lowball the risk.

People with poor numeracy considered graphics more helpful for understanding the risk information, which is consistent with previous findings about subjective numeracy.

In the substudy of unlabeled graphics, better numeracy was correlated with better accuracy in estimating the proportion in the graphic, and low-numeracy respondents gave higher mean estimates for all graphics than did high-numeracy ones. A number of other researchers have found that interpretation of risk graphics such as survival curves improves with either previous education or targeted training. However, previous studies of simple tasks such as estimating proportions have been conducted in educationally homogeneous groups or have not assessed educational level or numeracy as a covariate. It is possible that the computational skills measures by the Lipkus questions contribute to the ability to interpret graphics. Alternately, the ability to estimate quantities visually could itself contribute to the development of the computational skills.

As previously noted, low-numeracy respondents typically reported much higher risk perceptions and stronger intentions to take preventive action than high-numeracy respondents did. However, the interactive search graphics nearly eliminated these differences between high- and low-numeracy respondents by reducing perceived risk among the less numerate and also slightly increasing perceived risk among the more numerate. This effect suggests that graphics such as these could improve communication by reducing differences between the way that numerate healthcare professionals and less-numerate patients perceive risks. It also suggests that graphics such as these might be particularly effective when the goal is to reduce overactions to risks among the less numerate, such as situations in which high perceived risks lead people to reject health promotion communications (due to fear).

Some themes that can tie the studies detailed in this report are as follows: It is certainly possible and potentially beneficial to tailor graphics for people with different levels of expertise or skills. The current data add to the literature showing that numeracy varies widely even among well-educated people and strongly affects the interpretation of graphics. The interactive search graphics developed in this project, however, sharply reduced the differences in interpretation and decisions between high- and low-numeracy readers. These or other types of graphics could be used to compensate for low numeracy, thus helping people perform at a higher level. A particularly promising and under-research area is the topic of tailoring risk graphics, other types of data graphics, and numerical information in general to compensate for low numeracy. A second theme is the importance of designing graphics to support specific goals (i.e., information, education, persuasion, or decisions). Design decisions can have unintended consequences on these goals. Sometimes, the goal is to persuade people to avoid hazards and adopt health lifestyles. In other cases, such as recruiting participants for medical research, persuasion is unacceptable. These considerations in light of our findings warrant careful consideration.

6. LIST OF PUBLICATIONS AND PRODUCTS

1. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs for communicating health risks: A systematic review. *Journal of the American Medical Informatics Association* 2006; 13(6): 608-618 (electronic preprint ahead of print, August 2006).

2. Ancker JS. Health literacy and health numeracy [panel presentation]. AMIA 2006 Annual Symposium, Hilton Towers, Wash. DC, Nov. 11-15, 2006.

3. Ancker JS, Kukafka R, Weber EU. Risk and experience: Effects of experiential learning and patient characteristics in interpretation of dynamic risk graphics [poster]. *Proceedings/AMIA Annual Fall Symposium* 2006.

4. Ancker JS, Kukafka R. A combined qualitative method for testing an interactive risk communication tool. *Proceedings/AMIA Annual Fall Symposium* 2007: 16-20.

5. Ancker JS, Chan C, Kukafka R. Interactive graphics to demonstrate health risks: formative development and qualitative evaluation. J Health Commun. 2009 Jul-Aug;14(5):461-75.

6. Ancker JS, Weber EU, Kukafka R. Effects of interactive computer graphics on risk perceptions and decisions. Journal of Medical Decision Making. In Press.

LITERATURE CITED

1. Becker MH. The health belief model and personal health behavior. *Health Educ Monogr.* 1974(2):324-508.
2. Weinstein ND. The precaution adoption process. *Health Psychol.* 1988;7:355-386.
3. Witte K. Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs.* 1992;59:329-349.
4. Brewer NT, Gibbons FX, Gerrard M, Chapman GB, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychol.* 2007;26(2):136-145.
5. 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.
6. Woolf S, Chan ECY, Harris R, et al. Promoting informed choice: transforming health care to dispense knowledge for decision making. *Ann Intern Med.* 2005;143(4):293-300.
7. Hibbard JH, Peters E. Supporting informed consumer health care decisions: Data presentation approaches that facilitate the use of information in choice. *Annu Rev Public Health.* 2003;24:413-433.
8. Fuller R, Dudley N, Blacktop J. How informed is consent? Understanding of pictorial and verbal probability information by medical inpatients. *Postgrad Med.* 2002;78(543-544).
9. Ancker JS, Kaufman DR. Rethinking health numeracy: A multidisciplinary literature review. *J Am Med Inform Assoc.* 2007;14(6):713-721.
10. Schwartz L, Woloshin S, Black W, Welch H. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127(11):966-972.
11. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21:37-44.
12. Gurmankin AD, Baron J, Armstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication *Med Decis Making.* 2004;24(3):265-271.
13. Weinstein ND, Atwood K, Puleo E, Fletcher R, Colditz G, Emmons KM. Colon cancer: Risk perceptions and risk communication. *Journal of Health Communication.* 2004;9 (1):53-65.
14. Grimes DA, Snively GR. Patients' understanding of medical risks: Implications for genetic counseling. *Obstet Gynecol.* 1999;93:910-914.
15. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: A systematic review. *J Am Med Inform Assoc.* 2006;13(6):608-618.
16. Lipkus IM, Hollands JG. The visual communication of risk. *JNCI Monographs.* 1999;25:149-163.
17. Armstrong K, Schwartz JS, Fitzgerald G, Putt m, Ubel PA. Effect of framing as gain versus loss on understanding and hypothetical treatment choices: survival and mortality curves. *Med Decis Making.* 2002;22:76-83.
18. Armstrong K, Fitzgerald G, Schwartz JS, Ubel PA. Using survival curve comparisons to inform patient decision making: Can a practice exercise improve understanding? *J Gen Intern Med.* 2001(16):482-485.
19. Stone ER, Sieck WR, Bull BE, Yates JF, Parks SC, Rush CJ. Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. *Organ Behav Hum Decis Process.* 2003;90(1):19-36.
20. Schirillo JA, Stone ER. The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Anal.* 2005;25(3):555-566.
21. Schapira MM, Nattinger AB, McAuliffe TL. The influence of graphic format on breast cancer risk communication. *Journal of Health Communication.* 2006;11(6):569-582.
22. Elting L, Martin C, Cantor S, Rubenstein E. Influence of data display formats on physician investigators' decisions to stop clinical trials. *Br Med J.* 1999;318:1527-1531.

23. Schapira MM, Nattinger AB, McHorney CA. Frequency or probability? A qualitative study of risk communication formats used in health care. *Med Decis Making.* 2001;21:459-467.
24. Petty R, Wegener D. The elaboration likelihood model: Current status and controversies. In: Chaiken S, Trope Y, eds. *Dual-Process theories in Social Psychology*. New York: Guildford Press; 1999.
25. Emmons KM, Wong M, Puleo E, Weinstein ND, Fletcher R, Colditz G. Tailored computer-based cancer risk communication: Correcting colorectal cancer risk perception. *Journal of Health Communication.* 2004;9:127-141.
26. Natter HM, Berry DC. Effects of active information processing on the understanding of risk information. *Applied Cognitive Psychology.* 2005;19:123-135.
27. Hertwig R, Barron G, Weber EU, Erev I. Decisions from experience and the effect of rare events in risky choice. *Psychological Science.* 2004;15(8):534-539.
28. Weber EU, Shafir S, Blais AR. Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychol Rev.* 2004;111(2):430-445.
29. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica.* 1979;47:263-291.
30. Ancker JS, Kukafka R. A combined qualitative method for testing an interactive risk communication tool. *Proc AMIA Annu Fall Symp.* 2007:16-20.
31. Rosson MB, Carroll JM. *Usability engineering: scenario-based development of human-computer interaction* San Francisco: Morgan Kaufmann; 2002.
32. Nielsen J. *Usability Engineering*. Cambridge (MA): Academic Press; 1993.
33. Strauss A, Corbin J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks, CA: Sage; 1998.
34. Royak-Schaler R, Blocker D, Yali A, Bynoe M, Briant K, Smith S. Breast and colorectal cancer risk communication approaches with low-income African-American and Hispanic women: implications for healthcare providers. *J Natl Med Assoc.* 2004;96(5):598-608.
35. National Cholesterol Education Program. *Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III): Executive Summary* National Heart, Lung, and Blood Institute; National Institutes of Health; 2001.
36. Weinstein ND, Kwitel A, McCaul KD, Magnan RE, Gerrard M, Gibbons FX. Risk perceptions: assessment and relationship to influenza vaccination. *Health Psychol.* 2007;26(2):146-151.
37. Baker D, Williams M, Parker R, Gazmararian J, Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999;38(1):33-42.
38. Nurss J, Parker R, Williams M, Baker D. *TOFHLA test of functional health literacy in adults*. Show Camp, NC: Peppercorn Books and Press; 2001.
39. Peters E, Vastfjall D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. *Psychological Science.* 2006;17(5):407-413.
40. Weinstein ND, Sandman PM, Hallman WK. Testing a visual display to explain small probabilities. *Risk Anal.* 1994;14(6):895-896.