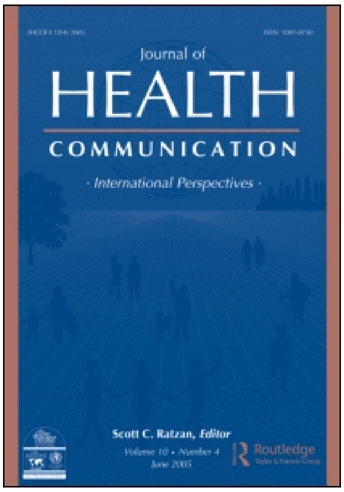

Journal of Health Communication

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713666566

Presentation Format Affects Comprehension and Risk Assessment: The Case of Prenatal Screening

Talya Miron-Shatz [a]; Yaniv Hanoch [b]; Dana Graef [c]; Michal Sagi [d]
[a] Princeton University, Princeton, New Jersey, USA [b] University of Plymouth, Plymouth, United Kingdom [c] Yale University, New Haven, Connecticut, USA [d] Hebrew University, Jerusalem, Israel

Online Publication Date: 01 July 2009

Journal of Health Communication, 14:439–450, 2009

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730903032986

Routledge
Taylor & Francis Group

Presentation Format Affects Comprehension and Risk Assessment: The Case of Prenatal Screening

TALYA MIRON-SHATZ

Princeton University, Princeton, New Jersey, USA

YANIV HANOCH

University of Plymouth, Plymouth, United Kingdom

DANA GRAEF

Yale University, New Haven, Connecticut, USA

MICHAL SAGI

Hebrew University, Jerusalem, Israel

*We studied the effects of presentation formats (frequency, 1-in-*N*, and visual) and numeracy level on students' understanding of prenatal screening results, as well as their risk assessment for having a fetus with Down syndrome. Frequency format (vs. 1-in-*N *and visual formats) improved participants' ability to accurately assess the chances of having a fetus with Down syndrome, and was associated with lower risk estimates. High numeracy levels were associated with a better ability to judge risk likelihood. For individuals of low numeracy levels, however, the frequency format significantly facilitated accurate understanding of probability information. This suggests that presenting information in frequency format is particularly beneficial for certain populations.*

In January 2007, The American College of Obstetricians and Gynecologists published new guidelines for prenatal screening, advising all pregnant women, not just those 35 or older, to undergo screening (American College of Obstetricians and Gynecologists, 2007). To facilitate the decision-making process, the American College of Obstetricians and Gynecologists recommends that "information about the detection and false-positive rates, advantages, disadvantages, and limitations, as well as the risk and benefits of diagnostic procedure should be available to

We thank Hilary Louise Bekker, Wolfgang Gaissmaier, Mirta Galesic, Shimon Saphire-Bernstein, Anita Todd, and two anonymous reviewers for comments on earlier drafts, and Amy Krilla for excellent research assistance. The first and second authors contributed equally to the preparation of the article. The study was supported by the Center for Health and Wellbeing, Princeton University.

Address correspondence to Yaniv Hanoch, University of Plymouth, School of Psychology, Drake Circus, Plymouth PL4 8AA, United Kingdom. E-mail: yaniv.hanoch@plymouth.ac.uk

patients so that they can make informed decisions'' (2007, p. 219). Providing information is an essential component of shared decision making, though it is unclear how health care providers should present prenatal screening information (Gates, 2004).

The new guidelines highlight the need to communicate probabilistic information to women effectively (Hofman et al., 1993). A number of barriers, however, can hamper the process. In his seminal work, Eddy (1982) showed that various health care professionals—doctors, nurses, and midwives—have difficulties understanding, communicating, and correctly solving statistical problems. Similarly, Reyna (2001) argued that health care professionals ''experience predictable difficulties in interpreting probabilities associated with genetic testing'' (p. 2406; see also Gigerenzer, 2002). Yet this is some of the very information that the American College of Obstetricians and Gynecologists recommends that patients be given.

Presentation Format

Risk information is most typically presented in probabilistic terms, possibly undermining people's ability to comprehend it (Grimes & Snively, 1999; Statham & Green, 1993). For example, Ekwo, Seals, Kim, Williamson, and Hanson (1985) examined pregnant women's perceived risk of having a child with Down syndrome and the association between their risk perception and their decision to undergo prenatal testing (amniocentesis). The majority of the women misunderstood their risk levels even though they received that information in both verbal (e.g., very low) and probability formats (e.g., 1 in 800).

More recently, Dewhurst and colleagues (2007) compared the ability of undergraduate students, genetic counseling graduate students, and genetic counselors to identify the correct probabilities in four typical problems encountered in genetic counseling (birth order, hemophilia, breast cancer, and Down syndrome). Only 9.5% of the undergraduates, 42% of the graduate students, and 46% of the genetic counselors answered the Down syndrome question correctly (the correct answer was 0.18%). The others provided incorrect answers that ranged from 0 to 900%.

In an impressive study with over 16,000 participants, Cutie, Weinstein, Emmons, and Colditz (2008; see also Woloshin, Schwartz, Byram, Fischhoff, & Welch, 2000) presented participants with a number of hypothetical medical questions in one of three formats: percentage, frequencies, and 1-in-*N*. After reading a short question, participants were asked to perform a number of mathematical operations (e.g., compare between two probabilities of having cancer) that often are encountered in the discussion of health risk information. As with previous studies, participants' ability to solve health risk information was linked directly to the presentation format: participants who encountered risk probability information in either percentage and frequency format did better than those who faced information in the 1-in-*N* format.

If the 1-in-*N* format is difficult to grasp (Cutie et al., 2008), what are other ways to present statistical or risk-related information that can facilitate understanding? Two such alternatives are visual and natural frequency formats. Research by Gigerenzer and Hoffrage (1995; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000) has shown that doctors, medical students, and laymen can improve their statistical performance if information is presented in a natural frequency format. Following

Gigerenzer and Hoffrage's work, Bramwell, West, and Salmon (2006) compared the ability of pregnant women and their companions, midwives, and obstetricians to interpret prenatal screening results correctly, such as the chance of a fetus having Down syndrome, as a function of the presentation format used (probability expressed as a percentage vs. a frequency). While only 14% of the participants provided the correct answer when given the probability format, those given the frequency format did significantly better. This effect, however, was present only among obstetricians, whose correct answer rates were only 5% for the probability format and 63% for the frequency format.

While the frequency format seems to be a superior alternative to the percentage or standard 1-in-*N* formats, researchers (Ancker, Senathirajah, Kukafka, & Starren, 2006; Fagerlin, Wang, & Ubel, 2005; Lipkus, 2006) have argued that a visual presentation can aid participants' ability to interpret statistical information. In one study, Fagerlin and colleagues presented participants with two possible treatments for angina—balloon angioplasty or bypass surgery. The participants who received pictorial rather than verbal information were better able to solve statistical information and to reduce the influence of anecdotal information when deciding on a treatment.

Numeracy

Changing the presentation format can facilitate people's understanding of statistical information. Numeracy, usually defined as the ability to understand basic mathematical concepts (Schwartz, Woloshin, Black, & Welch, 1997), could be another contributing factor, as it has been associated with the way people perceive risk (Lipkus, Samsa, & Rimer, 2001). Lipkus and colleagues, who developed a numeracy scale, found that even "highly educated" people could not solve basic probability problems such as distinguishing magnitudes of 2, or converting a chance of 1 in 1,000 into a percentage. Examining 500 female veterans, Schwartz and colleagues (1997) found a strong correlation between numeracy levels and the ability to estimate correctly the benefits of mammography screening.

Low numeracy rates, however, are not endemic to lay people. When six numeracy questions were administered to doctors, nurses, doctorate faculty, and medical students, Estrada, Barnes, Collins, and Byrd (1999) found that only slightly more than half of the people in this highly educated sample were able to answer all six questions correctly.

In a series of studies, Peters and colleagues (2006) investigated the relationship among numeracy, decision quality, and affective reactions. For example, in Study 2 the authors examined the relationship between numeracy and risk representation on a task estimating the likelihood that a mental patient would commit a violent act within a year after discharge. While presentation format (percentage vs. frequency) had no effect on risk rating for the high-numeracy group, the low-numeracy group rated the risk as much lower when information was presented as a percentage (e.g., 10%).

In light of the above research, we tested a number of hypotheses. Our main hypothesis was that presentation format affects comprehension and risk assessment. Specifically, we predict that frequency and visual (vs. 1-in-*N*) formats will improve participants' comprehension of statistical information, with the frequency format eliciting the greatest improvement. A second hypothesis is that participants with high numeracy levels will have a better understanding of the statistical information.

Method

Participants

The participants were 241 students from a private university in the Northeastern United States. They were recruited to complete an online survey about medical decision making via e-mails and postings around campus. As an incentive for participation, participants were entered in a raffle for three $100 cash rewards. The mean age of our participants was 21.14 ($SD = 3.58$), and all of them reported that they had no children and no prior experience with the prenatal screening test. Most participants (61%) were women (mean age 20.96 years, $SD = 3.14$), and 39% were men (mean age 21.42 years, $SD = 4.18$).

Design, Procedure, and Instrument

After agreeing to participate in the study, participants read a short paragraph with background information about genetic screening that was followed by a "letter to an expectant mother," depicting a realistic prenatal screening result for the seventeenth week of pregnancy. The letter contained reports of the levels of three hormones (unconjugated estriol, alpha-fetoprotein, and human chorionic gonadotropin) in the maternal serum, and all participants were informed that "The combination of the three results is abnormal for the 17th week of pregnancy." Previous research indicates that labeling genetic screening results negatively, as "abnormal," is associated with increased risk perceptions and more negative emotional reactions (Zikmund-Fisher, Fagerlin, Keeton, & Ubel, 2007). We used this phrasing to add to the ecological validity of the study as it closely mirrored the language in letters sent to women by the Hadassah Medical Center in Israel following prenatal screening tests. Furthermore, since all participants received results with the same negative phrasing, any differences between groups can be attributed only to differences in the presentation format, and not to the negative phrasing per se.

After reading the raw results, participants encountered information communicating the probability that the woman was carrying a fetus with Down syndrome. The format in which this information was presented served as the key manipulation for the study. Participants were randomly assigned to one of three conditions (for a similar design with regard to women's breast cancer see Schapira, Nattinger, & McHorney, 2001). For participants in the first condition, the information was presented in a 1-in-*N* format (see Cutie et al., 2008; Denes-Raj, Epstein, & Cole, 1995): "The probability of giving birth to a baby with Down syndrome for a woman with normal results is 1:724. The probability of giving birth to a baby with Down syndrome for a woman with the abnormal result above is 1:181." In the second condition, the information was presented in a *frequency* format: "One out of every 724 fetuses of women your age will be diagnosed with Down syndrome. One out of every 181 fetuses of women your age with test results identical to yours will be diagnosed with Down syndrome." In the third condition, the information was presented in a *visual* format (see Ancker et al., 2006; Feldman-Stewart et al., 2007; Lipkus, 2006): "The white circle in the picture represents the only fetus with Down syndrome out of all the fetuses of women your age who have not yet been screened. There are 724 circles in the picture." Following this text was an image of 723 black circles and one white circle placed toward the middle. Below this image, additional text told the participants, "The white circle in the picture below represents the only fetus with

Down syndrome out of all the fetuses of women whose screening results were identical to yours. There are 181 circles." This text was followed by an image of 180 black circles and one white circle. In all conditions, the letter concluded by saying that an amniocentesis test was recommended "given the abnormal results and the elevated risk for Down syndrome," and the expectant mother was invited to come to the clinic for a genetic counseling session.

After reading the letter, participants in all conditions responded to questions assessing their comprehension of the probability that the woman was carrying a fetus with Down syndrome. Following this, participants' numeracy skills were assessed and they were asked to report their age, gender, whether they had any children, and whether they had ever undergone a maternal serum screening test.

Measures

Comprehension. Participants' comprehension of the probability that the woman was carrying a fetus with Down syndrome was assessed by asking "After reading the previous letter, what do you think the probability is that the woman is carrying a fetus with Down syndrome?" The response was open-ended, meaning that participants could enter the probability in whatever format they wished. Given that the correct response was a fraction of a percentage point (the probability was 1/181, or 0.55%), the use of the open-response format created the possibility that a simple comparison of means would suffer from artifacts due to a highly skewed distribution (if, for example, a few participants answered 50%, a response that is nearly 100 times greater than the correct response). To account for this potential problem, we derived two dependent measures from this response: *Mean comprehension* consisted of the responses recoded in a uniform numeric format. *Correct comprehension* was a dummy variable calculated to indicate whether the participant had provided a correct answer. A response was scored as correct if the reported probability was between 0.5% and 0.6%, or if the participant said "less than 1%."[1]

Risk Assessment. Immediately following the comprehension question, participants were asked to provide an assessment of the risk that the woman was carrying a baby with Down syndrome by selecting one of four answer choices (*low, medium, high*, or *I don't know*). In order to calculate mean risk assessments, these responses were assigned ordinal values as follows: 1 = *low*; 2 = *medium*; 3 = *high*. Participants who selected the "I don't know" response were excluded from the analysis for this variable.

Clarity. Participants rated how easy it would be for the woman who received the letter to determine whether a 1:181 chance of having a fetus with Down syndrome is a good or a bad result (see Hsee & Leclerc, 1998; Zikmund-Fisher, Fagerlin, Roberts, Derry, & Ubel, 2008), using a Likert-like rating scale of 0 (*not clear at all*) to 6 (*very clear*).

[1]Responses of "1/181," "1:181," "0.55%," and "one out of one hundred and eight-one" all would be considered correct. In addition, participants who responded "less than 1%" (or some variation of that statement) were considered correct, but note that this does not apply to participants who gave a *specific* numeric response outside the 0.5%–0.6% range that happened to be less than 1%.

Numeracy. To assess mathematical ability, participants filled out an 11-item numeracy scale originally developed by Lipkus and colleagues (2001) and revised by Peters et al. (2006). Participants' numeracy scores were assessed by the number of correct responses (maximum 11).

Recall. At the end of the survey, participants once again were asked what the probability was that the woman who had received the letter was carrying a fetus with Down syndrome. The question read, "Recalling the letter you read in the beginning of the survey, what do you think the chance is that the woman who received the letter has a fetus with Down syndrome?" The participants were required to select one of four responses (*no chance, 1% or less, between 2 and 3%*, and *4% or above*). As with comprehension, two dependent measures were derived from the response to the recall item: *Mean recall* was calculated by assigning the following values to the response options: 0 = *no chance*; 1 = 1% *or less*; 2 = *between* 2 *and* 3%, and 3 = 4% *or above*. *Correct recall* was a dummy variable calculated to indicate whether the participant had provided a correct answer, less than 1%.[2]

Note that due to the conceptual similarity between the recall measure and the earlier measures of comprehension and risk assessment, the analysis of recall appears immediately after the analysis of risk assessment in the results section. The recall question, however, appeared later in the survey. Here we maintain the precise chronological order of question presentation, but in the results section, we present the analyses in a manner more conducive to conceptual coherence.

Results

Gender

We checked for gender effects before conducting any analyses. There were no significant differences between by gender for the comprehension and risk assessment questions. While there was a significant effect of gender on the recall question, $F(1,235) = 4.63$, $p < .05$, $\eta_p^2 = .02$, the interaction with presentation format was not significant, $F(2,235) = .11$, *ns*. Therefore, we conducted the analyses of the effects of presentation format on these variables without controlling for genders.

Comprehension

The frequency presentation format was associated with a lower estimate of probability in the open-response question. The open-response probability estimates were influenced by presentation format, $F(2,227) = 7.54$, $p < .001$, $\eta_p^2 = .06$. A post-hoc analysis using Fisher's least significant difference (LSD) method showed that the estimates of the woman's risk were significantly higher in the 1-in-N format than they were in both the visual and the frequency format conditions, $p < .001$ for both (see Table 1).

[2]As with the comprehension variable, participant responses were translated into a numerical percentage, so responses of "1/181," "1:181," "0.55%," and "one out of one hundred and eight-one" all would be considered correct. In addition, participants who responded "less than 1%" (or some variation of that statement) were considered correct, but note that this does not apply to participants who gave a *specific* numeric response outside the 0.5%–0.6% range that happened to be less than 1%.

Table 1. Mean response by condition

	Mean response (SD)			
Dependent variable	1-in-*N*	Visual	Frequency	*F*
Mean comprehension	0.16^ (0.21)	0.07+ (0.14)	0.06+ (0.17)	7.54***
Correct comprehension[a]	0.36^ (0.49)	0.55+ (0.50)	0.67+ (0.47)	7.42***
Risk assessment[b]	1.72^ (0.67)	1.43+ (0.54)	1.42+ (0.65)	5.49**
Mean recall[c]	1.27^ (0.58)	1.28^ (0.62)	1.08+ (0.32)	3.58*
Correct recall[a]	0.79^ (0.41)	0.81^ (0.40)	0.93+ (0.25)	3.50*
Clarity	2.61^ (1.51)	2.07+ (1.44)	2.72^ (1.47)	4.73**
Numeracy	10.19^ (1.11)	10.26^ (1.28)	10.27^ (1.09)	0.09 *ns*

Note. The *F* value is the result of a univariate analysis of variance test for differences in the mean level of the dependent variable in the corresponding row, across the three conditions. The ^ and + symbols identify significant pairwise comparisons. If two groups are not significantly different for a given variable, they receive the same sign. Group means flagged with a different symbol are significantly different from the other groups.

[a]The means for this variable also can be read as a percentage, as answers were scored 1 if correct and 0 if incorrect. Thus, for example, 36% of participants in the probabilistic condition gave the correct answer.

[b]Answers were scored 1 for "low," 2 for "medium," and 3 for "high." Responses of "I don't know" were not included in the analysis.

[c]Answers were scored 0 for "no chance,"1 for "less than 1%," 2 for "2–3%," and 3 for "4% or greater." As no participants select the "no chance" option, a lower mean response for a group indicates that the group as a whole was more correct, relatively speaking.

*$p < .05$. **$p < .01$. ***$p < .001$.

There was also a main effect of presentation format on the correct comprehension variable, $F(2,227) = 7.42$, $p = .001$, $\eta_p^2 = .06$. A post-hoc analysis (Fisher's LSD) indicated that the proportion of participants who gave a correct answer to the comprehension question was significantly smaller in the 1-in-*N* format condition than the proportions of participants who gave the correct response in both the visual and the frequency format conditions, $p < .05$ and $p < .001$, respectively. The high degree of similarity between the results of the analyses for mean comprehension and correct comprehension suggest that our fear of an artifact resulting from skewed distributions was unfounded in this case.

Risk Assessment

Presentation format had a significant effect on participants' risk assessments, $F(2,226) = 5.49$, $p < .01$, $\eta_p^2 = .05$. A post-hoc analysis (Fisher's LSD) showed that the risk assessments in the 1-in-*N* format condition were significantly higher than the assessments in both the visual and the frequency format conditions, $p < .01$ for both.

Recall

Presentation format also had a significant effect on the mean recall variable, $F(2,238) = 3.58$, $p < .05$, $\eta_p^2 = .03$. A post-hoc analysis (Fisher's LSD) showed that mean recalled probability of Down syndrome in the frequency format condition was significantly lower, hence more accurate, than mean recall in both the 1-in-N and the visual format conditions, $p < .05$ for both. A similar effect was found for correct recall, which also differed significantly by condition, $F(2,238) = 3.50$, $p < .05$. A post-hoc analysis (Fisher's LSD) showed that the proportion of participants who answered correctly to the recall question in the frequency format condition was higher than the proportion of participants who got the recall question right in both the probability and the visual format conditions, $p < .05$ for both.

Clarity

Presentation format also influenced the perceived clarity of the results, $F(2,238) = 4.73$, $p < .01$, $\eta_p^2 = .04$. The 1-in-N format elicited a mean clarity rating that was not significantly different from the mean clarity rating for the frequency format. The visual presentation format was perceived as being less clear than the other two formats, and a post-hoc analysis (Fisher's LSD) revealed that the ratings for this format were significant lower than the ratings for both the 1-in-N and the frequency formats, $p < .05$ and $p < .01$, respectively.

Numeracy

To analyze the effect of numeracy on probability comprehension and risk assessment, we conducted a median split to form "high" and "low" numeracy groups. As only 12.4% of our sample had a numeracy score of 8 or under, we used a median slit rather than a cut-off point of 11 to 9 and 8 or under. The fact that the sample was relatively numerate, however, only serves to increase our concerns that low numeracy (in absolute and relative terms) is associated with decreased understanding and with heightened risk estimates. The "high" numeracy group consisted of the 54.8% of participants who reached the maximal numeracy rating by answering all 11 numeracy questions correctly. The rest of the participants were in the "low" numeracy group. Overall, 28.2% of the participants received a numeracy score of 10, with only 17% getting nine or fewer questions correct. The mean numeracy score for all participants was 10.24 ($SD = 1.17$).

Numeracy did not have a main effect on mean comprehension, $F(1,228) = 1.02$, *ns*, nor was there a difference between the two numeracy groups in the proportion of participants who gave the correct answer to the comprehension question, $F(1,228) = 0.001$, *ns*. In addition, two-way ANOVAs [numeracy X presentation format] did not find significant interaction effects on either mean comprehension, $F(2,224) = 0.05$, or on correct comprehension, $F(2,224) = 0.216$, both *ns*.

Numeracy did have an effect on risk assessment, $F(1,227) = 4.93$, $p < .05$, $\eta_p^2 = .02$. The high-numeracy group assessed the risk to be lower ($M = 1.44$, $SD = 0.61$) than did the low-numeracy group ($M = 1.62$, $SD = 0.64$). Numeracy also had a main effect on mean recall, $F(1,239) = 12.86$, $p < .001$, $\eta_p^2 = .05$, with the high-numeracy group having significantly lower recall scores ($M = 1.11$, $SD = .38$) than the low-numeracy group ($M = 1.35$, $SD = 0.66$). Consistent with this finding, the proportion of participants in the high-numeracy group who answered the recall question correctly ($M = 0.92$, $SD = .28$) was significantly greater than the proportion of low-numeracy participants who answered the recall question correctly ($M = 0.75$, $SD = 0.43$), $F(1,239) = 12.68$, $p < .001$, $\eta_p^2 = .05$.

Table 2. Frequencies of correct responses to recall by presentation format and numeracy

Presentation format	*n* correct/*n* per cell (% correct)	
	Low numeracy	High numeracy
1-in-*N*	26/36 (72.2%)	35/41 (85.4%)
Visual	26/40 (65.0%)	46/49 (93.9%)
Frequency	30/33 (90.9%)	40/42 (95.2%)

Note. Participants were given four choices: *No chance, less than* 1%, 2–3%, 4% *or above*. The correct response was *less than* 1%. As reported in the text, there were main effects for presentation format, $F(2,235) = 4.04$, $p < .05$, and for the numeracy median split, $F(1,235) = 11.52$, $p < .001$. The interaction was not significant, $F(2,235) = 2.57$, $p = .08$.

Table 2 displays the percentages of participants who gave the correct response to the recall question. The results are broken down by numeracy and presentation format. Below we discuss the implications for the low-numeracy group, as we anticipated that presentation format might make a larger difference for this group than for the high-numeracy group, as numeracy entails being able to interpret and shift between presentation formats.

Low Numeracy

While an ANOVA did not reveal a statistically significant main effect of numeracy on the comprehension question, analyses within the low-numeracy group show that this group suffers when faced with the 1-in-*N* presentation format. One-way ANOVAs within the low-numeracy group revealed a statistically significant effect of presentation format on mean comprehension, $F(2,99) = 3.43$, $p < .05$, $\eta_p^2 = .07$. A post-hoc analysis (Fisher's LSD) showed that the mean response in the 1-in-*N* format condition ($M = 0.18$, $SD = 0.22$) was significantly higher than the mean responses in both the frequency ($M = 0.09$, $SD = 0.20$) and the visual format conditions ($M = 0.07$, $SD = 0.12$), $p < .05$ for both. Presentation format did not have a statistically significant effect on the proportion of the participants who got the comprehension question correct in the low-numeracy group, $F(2,99) = 3.05$, *ns.* A post-hoc analysis (Fisher's LSD) did show, however, that the proportion of correct responses in the frequency format condition ($M = 0.66$, $SD = 0.48$) was significantly higher than the proportion of correct responses in the 1-in-*N* format condition ($M = 0.36$, $SD = 0.49$), $p < .05$.

One-way ANOVAs within the low-numeracy group also demonstrated a significant main effect of presentation format on risk assessments, $F(2,100) = 3.83$, $p < .05$, $\eta_p^2 = .07$, and on the recall question, $F(2, 106) = 3.58$, $p < .05$, $\eta_p^2 = .06$. In both cases the frequency presentation format was associated with a lower, more accurate estimation of risk. A post-hoc analysis (Fisher's LSD) of the risk assessments showed that the mean assessment was significantly lower for the frequency format ($M = 1.47$, $SD = 0.57$) than it was for the 1-in-*N* format ($M = 1.86$, $SD = 0.69$), $p < .05$. A similar test for the recall question showed that the mean for the frequency format ($M = 2.12$, $SD = 0.415$) was significantly lower than the mean for the visual format ($M = 2.53$, $SD = 0.78$), $p < .01$. As displayed in Table 2, the percentage of low-numeracy participants who gave the correct response to the recall question was only 65% in the visual format, but 91% of the participants gave the correct response in the frequency format.

Discussion

Some women who undergo prenatal screening tests, such as those at the Hadassah Medical Center in Israel, whose letter we used as a model, receive their results by mail. These letters convey information in a 1-in-*N* format. Our research findings seriously question the merit of using the 1-in-*N* format to inform women about their prenatal tests results.

As with earlier studies (e.g., Gigerenzer & Hoffrage, 1995), our work clearly illustrates the advantages of using the frequency format to convey statistical information. In support of our hypothesis, the frequency format (vs. the 1-in-*N* and visual formats) helped participants reach a more moderate and accurate estimation of risk. These findings are important for at least two reasons. First, as numeracy levels did not impact participants' responses to the open-response probability estimation, presentation format might be even more crucial in facilitating people's understanding of statistical information. It also should be noted that our findings stand in contrast to earlier work that shows a relationship between numeracy levels and an ability to solve mathematical problems (Lipkus et al., 2001). Given our well-educated and highly numerate participant pool, the relatively low rate of correct open responses was surprising. It also could be due to the perceived lack of clarity across all formats. Furthermore, the fact that the probability format was considered to be almost as clear as the frequency format indicates that participants are not necessarily good judges of how well a format conveys information.

Second, the task in our experiment was less difficult than those used in previous studies (e.g., Hoffrage et al., 2000), as participants were asked only to interpret the meaning of 1:181, rather than solve conditional probability questions based on Bayes's theorem. It is striking that, regardless of numeracy, half of the participants could not solve the simple probabilistic problem we presented to them. Finally, our data show that while numeracy facilitated lower numeric risk estimations (e.g., less than 1%), it interacted with presentation format. Only in the natural frequency format condition did the majority of the participants provide accurate risk estimates.

Furthermore, when asked to provide their own risk estimation, nearly half of our participants used the frequency format. Our results provide further support to Gigerenzer and Hoffrage's (1995) and Cutie and colleagues' (2008) notion that the frequency representation is the natural way by which people process information.

The American College of Obstetricians and Gynecologists' new guidelines for prenatal screening are bound to affect many women, and, hence, we believe that their guidelines should be cognizant of the findings reported in this study. As argued by a number of researchers (Ancker et al., 2006; Cutie et al., 2008; Feldman-Stewart et al., 2007; Lipkus, 2006), the way in which health care professional present probability information profoundly can affect the understanding of health risk information, as well as behavioral intentions (e.g., Marteau et al., 1991).

Our study offers several insights into the way probabilistic information should be presented. We did, however, use a sample that does not necessarily represent the general population. College students are, by definition, more educated and hence have higher levels of literacy and numeracy. Moreover, we asked participants to think about a hypothetical person, and not themselves. Given that most of our participants were less than 22 years old (and almost 40% were males), many of them might not have thought about prenatal screening and might have had difficulties imagining how a woman in the scenario might feel. Our study results, therefore, might not generalize to women who undergo prenatal screening. It is important, therefore, that future studies focus on women as they undergo prenatal screening tests and examine their understanding of statistical information and their numeracy levels. Given that health care professionals themselves have difficulties interpreting and communicating statistical information, our results could play an important role in alleviating some of the confusion.

References

American College of Obstetricians and Gynecologists (ACOG). (2007). ACOG practice bulletin: Screening for fetal chromosomal abnormalities. *Obstetrics & Gynecology*, *109*, 217–227.

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*, 608–618.

Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: Experimental study. *British Medical Journal*, *333*, 284–289.

Cutie, C. L., Weinstein, N. D., Emmons, K., & Colditz, G. (2008). A test of numeric formats for communicating risk probabilities. *Medical Decision Making*, *28*, 377–384.

Denes-Raj, V., Epstein, S., & Cole, G. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, *21*, 1083–1092.

Dewhurst, M., Veach, P., Lampman, C., Petraitis, J., Kao, J., & LeRoy, B. (2007). Probability biases in genetic problem solving: A comparison of undergraduates, genetic counseling graduate students, and genetic counselors. *Journal of Genetic Counseling*, *16*, 157–170.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristic and biases* (pp. 249–267). Cambridge: Cambridge University Press.

Ekwo, E. E., Seals, B. F., Kim, J.-O., Williamson, R. A., & Hanson, J. W. (1985). Factors influencing maternal estimates of genetic risk. *American Journal of Medical Genetics*, *20*, 491–504.

Estrada, C., Barnes, V., Collins, C., & Byrd, J. C. (1999). Health literacy and numeracy. *Journal of the American Medical Association*, *282*, 527.

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, *25*, 398–405.

Feldman-Stewart, D., Brennenstuhl, S., McIssac, K., Austoker, J., Charvet A., Hewitson, P., Sepucha, K. R., & Whelan, T. (2007). A systematic review of information in decision aids. *Health Expectations*, *10*, 46–61.

Gates, E. A. (2004). Communicating risk in prenatal genetic testing. *Journal of Midwifery & Women's Health*, *49*, 220–227.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency format. *Psychological Review*, *102*, 684–704.

Grimes, D. A., & Snively, G. R. (1999). Patients' understanding of medical risks: Implications for genetic counseling. *Obstetrics & Gynecology*, *93*, 910–914.

Hsee, K., & Leclerc, F. (1998). Will products look more attractive when presented separately or together? *The Journal of Consumer Research*, *25*, 175–186.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.

Hofman, K. J., Tambor, E. S., Chase, G. A., Geller, G., Faden, R. R., & Holtzman, N. A. (1993). Physicians' knowledge of genetics and genetic tests. *Academic Medicine*, *68*, 625–632.

Lipkus, I. M. (2006). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*, 696–713.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44.

Marteau, T. M., Kidd, J., Cook, R., Michie, S., Johnston, M., Slack, J., et al. (1991). Perceived risk not actual risk predicts uptake of amniocentesis. *British Journal of Obstetrics & Gynaecology*, *98*, 282–286.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407–413.

Reyna, V. F. (2001). Genetic testing and medical decision making. *Archive of Internal Medicine*, *161*, 2406–2408.

Schapira, M. M., Nattinger, A. B. & McHorney, C. A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, *21*, 459–467.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*, 966–972.

Statham, H., & Green, J. (1993). Serum screening for Down's syndrome: Some women's experiences. *British Medical Journal*, *307*, 174–176.

Woloshin, S., Schwartz, L., Byram, S., Fischhoff, B., & Welch, H. (2000). A new scale for assessing perception of chance. A validation study. *Medical Decision Making*, *20*, 298–307.

Zikmund-Fisher, B. J., Fagerlin, A., Keeton, K., & Ubel, P. A. (2007). Does labeling prenatal screening test results as negative or positive affect a woman's responses? *American Journal of Obstetrics & Gynecology*, *197*, 528.e1–528.e6.

Zikmund-Fisher, B. J., Fagerlin, A., Roberts, T. R., Derry, H., & Ubel, P. A. (2008). Alternate methods of framing information about medication side effects: Incremental risk versus total risk of occurrence. *Journal of Health Communication*, *13*, 107–124.