Evaluation of risk communication in a mammography patient decision aid

Krystal A. Klein[*], **Lindsey Watson**, **Joan S. Ash**, and **Karen B. Eden**

Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, USA

Abstract

Objectives—We characterized patients' comprehension, memory, and impressions of risk communication messages in a patient decision aid (PtDA), Mammopad, and clarified perceived importance of numeric risk information in medical decision making.

Methods—Participants were 75 women in their forties with average risk factors for breast cancer. We used mixed methods, comprising a risk estimation problem administered within a pretest–posttest design, and semi-structured qualitative interviews with a subsample of 21 women.

Results—Participants' positive predictive value estimates of screening mammography improved after using Mammopad. Although risk information was only briefly memorable, through content analysis, we identified themes describing why participants value quantitative risk information, and obstacles to understanding. We describe ways the most complicated graphic was incompletely comprehended.

Conclusions—Comprehension of risk information following Mammopad use could be improved. Patients valued receiving numeric statistical information, particularly in pictograph format. Obstacles to understanding risk information, including potential for confusion between statistics, should be identified and mitigated in PtDA design.

Practice implications—Using simple pictographs accompanied by text, PtDAs may enhance a shared decision-making discussion. PtDA designers and providers should be aware of benefits and limitations of graphical risk presentations. Incorporating comprehension checks could help identify and correct misapprehensions of graphically presented statistics

Keywords

Patient decision aid; Risk communication; Mammography; Women's health

[*]Corresponding author at: Mail Code BICC,3181SW Sam Jackson Park Road, Portland, OR, 97239, USA. kleinkr@ohsu.edu, krystal@alumni.brandeis.edu (K.A. Klein).
watsolin@ohsu.edu (L. Watson), ash@ohsu.edu (J.S. Ash), edenk@ohsu.edu (K.B. Eden).

1. Introduction

Patient decision aids (PtDAs) are evidence-based tools that help patients engage in informed, shared decision making regarding complex health decisions, such as "preference-sensitive" [1] decisions—those where no "best" course of action exists across all patients. PtDAs differ from general educational materials by helping patients understand how their values relate to available options' attributes [2]. One touted benefit is that PtDAs allow more effective, balanced risk communication than typical clinical consultation [3]. Indeed, patients who use PtDAs along with typical care demonstrate knowledge and risk comprehension superior to control patients [4].

The present study partnered with a project evaluating changes in decision quality measures reported by patients after using Mammopad, a mobile device-optimized PtDA designed for patients to use prior to a clinic visit, either at home, or in a waiting room [5]. Mammopad helps average-risk women in their forties understand and consider the costs and benefits of screening mammography options, clarify their values in relation to those options, and empower them to discuss screening mammography with clinicians. Recently, recommendations for routine mammography screening for average-risk women in their forties have been questioned due to equivocal evidence of benefit, *e.g.*, from randomized trials investigating the impact of routine mammography on breast cancer mortality [6–8]. While some organizations maintain these recommendations, the United States Preventive Services Task Force (USPSTF) stated that the decision to begin mammography screening before age fifty "should be an individual one and take into account patient context, including the patient's values regarding specific benefits and harms" [9]—essentially deeming it a *preference-sensitive* decision. This standpoint is supported by researchers with expertise in patient-centered care [10,11].

The details of how to communicate risk to patients within a PtDA must be carefully considered. Currently, risk communication is as much an art as a science, despite a growing literature mapping the effects of numeric formats [12–18], viewer characteristics like statistical numeracy (ability to understand statistical information such as probabilities) [19–23], and chart types [24–27] on perception of risk levels, recall of statistics, and decision outcomes. We recorded reactions of rural-dwelling patients, in their own words, to numeric risk information in Mammopad. We aimed to characterize patients' comprehension, memory, and impressions of risk communication messages in Mammopad, and to clarify the role and perceived importance risk information has in medical decision-making.

2. Methods

We evaluated risk communication in Mammopad through mixed-methods triangulation using three approaches: (1) a quantitative pretest–posttest design where participants answered a word problem about the positive predictive value (PPV) of screening mammograms for women in their forties before and after using Mammopad; (2) a qualitative content analysis of what participants found valuable about numeric risk information; and (3) a qualitative content analysis of interpretation of risk communication diagrams, including misperceptions.

2.1. Participant recruitment and consent

2.1.1. Risk scenario participants—Participants in the Mammopad parent study–women in their forties at average risk of breast cancer according to the Breast Cancer Genetics Referral Screening Tool; BRST [29–31] (which included women with few or no risk factors) answered the risk scenario question described in Section 2.3.1 immediately before and after using Mammopad. The parent study's participants were recruited through chart review at three clinics identified through the Oregon Rural Practice-based Research Network (ORPRN) and eligibility screening phone calls. Rural clinics (i.e., in non-urbanized, medically underserved areas as defined by the State of Oregon) with low income patients were recruited to address concerns that these women may not be aware of their own breast cancer risk or have considered when to begin getting screening mammograms. Women in rural areas are less likely to have had a mammogram or to have an up-to-date mammogram [28]. Details of recruitment and participant flow into the parent study, including risk screening with the B-RST, were reported previously [5].

2.1.2. Interview participants—Early interview participants were a convenience subsample of Mammopad participants from two of the three clinics that volunteered for this follow-up study. After determining that all initial interviewees had previously had mammograms, we began recruiting participants without a previous mammogram. Because the Mammopad study completed enrollment before this study completed recruitment, we recalled some women who had participated previously. Participants were recruited to participate in a 30- to 40-min interview in exchange for a gift card. Recruitment ended when all women with no previous mammogram had either participated, declined, or were unreachable.

2.1.3. Consent—Participants consented separately for the parent Mammopad study and the semi-structured interview. Both protocols were approved by Oregon Health & Science University's Institutional Review Board.

2.2. Mammopad app and administration

Mammopad included three modules: a breast cancer informational module, a mammography informational module, and an interactive priority-setting module, which allowed users to prioritize harms and benefits of screening and identify questions and concerns to discuss with providers [5]. A summary report was presented to participants, and emailed to them, if requested. The numeric risk and probability graphics in Mammopad closely adhered to recent evidence-based guidelines for risk presentation [2,32]; they were refined through several rounds of discount usability testing of Mammopad.

In the parent study[5], a researcher loaded Mammopad for each participant on an Apple iPad mini 7.9-inch tablet computer. Participation occurred in a private clinic room, lasted about 30 min, and was observed at a distance by the researcher.

2.3. Data collection

2.3.1. Risk scenario phase—A risk scenario question assessed the participants' perception of the PPV of mammography (the breast cancer risk associated with abnormal

mammogram results). The question, shown in Box 1 , was posed immediately before and after using Mammopad. Participants responded using iPad mini's on-screen keyboard.

2.3.2. Interview phase—Semi-structured interviews were administered by the first author in private examination rooms at ORPRN-affiliated clinics (Appendix A contains the interview guide). The interviewer probed for recall of risk statistics (focusing on verbal explanations and numeric recall), and elicited participants' evaluations and explanations of what numbers were useful when consuming health information. Participants also reviewed Mammopad screenshots and discussed them with the interviewer. Screenshots explaining statistical information included a pair of breast cancer incidence graphics for women in their forties and fifties, (Fig. 1) and a mammography outcomes graphic (Fig. 2).

After the interview, participants answered a questionnaire evaluating their objective numeracy skill—the 3-min Berlin Numeracy Test [33,34] (reproduced in Appendix B) recommended for participants with low numeracy [35]. The Berlin Numeracy Tests are rigorously developed instruments, with robust discriminability, that strongly predict risk comprehension in everyday contexts [33]. By measuring our sample's command of statistical concepts relevant to risk understanding, we aimed to describe participants and confirm representation of women with different levels of numeracy.

Interviews were audio-recorded with participant consent, then transcribed by paid transcriptionists. We reviewed all transcripts for accuracy.

2.4. Analysis methods

2.4.1. Risk scenario phase—Descriptive and inductive statistics were completed using IBM SPSS Statistics 22 (IBM Corporation, Somers, NY, USA). We evaluated change in risk estimates between pre- and post-Mammopad using a Wilcoxon signed-rank significance test. Mann-Whitney U tests were employed to compare change scores across groups for two dichotomous pseudo-independent variables: previous mammogram experience, and knowing a friend or family member with breast cancer.

2.4.2. Interview phase—Participants were asked what numbers they remembered from Mammopad; responses were coded and tabulated. Outside of these quantitative descriptions, interview transcriptions were analyzed by a multidisciplinary mix of clinical research, informatics, computer science, and social science experts using conventional qualitative content analysis [36] to identify themes. Coding and analysis were completed using NVivo 10 Software for Mac (QSR International Pty., Ltd., Doncaster, Victoria, Australia). Initially, independent coding of three transcripts was undertaken by a primary coder (KAK) and one additional coder using an open coding framework, which allowed analysts to add new codes at any time. The coders then met to review transcript coding and consensually finalize an initial coding scheme. This scheme was used to code remaining transcripts, with each transcript being coded by the primary coder and one secondary coder. Review sessions were held after every two to three transcripts coded and resulted in consensus coding of each transcript recorded in NVivo; the coding schema was updated and redistributed after each session. After coding all transcripts, the analytic team independently critiqued the codes and came to consensus about refinements to the coding scheme from a top-down perspective.

Then, for each code in the scheme, the primary coder and one secondary coder reviewed coded references and drafted recommendations regarding changes to the schema. The analytic team then compared independent theme reports and jointly connected and organized codes/themes, resulting in a final theme list. Last, we selected quotations to illustrate themes. Trustworthiness [37], the qualitative research equivalent of validity, was established through use of standard methods to ensure reflexivity, multiple researchers analyzing the data, inclusion of an experienced qualitative researcher (JSA) on the project team, triangulation between quantitative and qualitative methods, recruiting from and testing at two different clinics, and bracketing quotations.

3. Results

Participant demographics are presented in Table 1. Berlin numeracy score was assessed for interview participants (except one, due to researcher error). Possible scores ranged from 0 (low) to 4 (high) with no partial credit. The sample mean score was 1.8 (sd = 1.15), and scores ranged from 0 to 4, confirming women of varying levels of numeracy were represented. Individual characteristics of interview participants appear in Appendix C.

3.1. Risk scenario phase

Unambiguous non-numeric responses were converted to numbers on a 0–100 scale (e.g., "Fifty-fifty" became 50; "twenty percent" became 20). Six of 75 participants were excluded for entering a non-numeric response, resulting in sixty-nine response pairs being analyzed.

After using Mammopad, the participants reduced their ratings of PPV of mammography from a median of 50 to a median of 6 (see Table 2). This change was in the direction of increased accuracy: based on Mammopad's risk graphics, the target answer was 2.0 or 3.0 (out of 100), depending on whether "pre-cancer" was classified as cancer. By a Wilcoxon signed-rank test, this reduction was statistically significant: $Z[69] = 5.721$, $p < .001$; there were 49 negative shifts, 4 positive shifts, and 16 ties. Mann-Whitney U tests found no significant differences in estimates between women with and without previous mammograms, or with and without a friend or family member with breast cancer (Table 3).

3.2. Interview phase

Twenty-one Mammopad participants were interviewed. Interview sessions immediately followed Mammopad sessions for twelve early-enrolled participants. Because the study lacked participants who had never had a mammogram when the parent study concluded enrollment, nine later-enrolled participants returned for the interview weeks or months after their initial sessions. Thus, we gathered both immediate and delayed impressions of Mammopad.

3.2.1. Themes: valuing numeric risk information—Three themes discovered in our analysis described reasons participants value numeric risk information. See quotations in Table 4.

3.2.1.1. Theme 1: valuing grounding in real-world groups: A frequently-mentioned theme was grounding in real-world groups: participants liked seeing an incidence probability

framed in terms of a number of people who might be in a room together in the real world (termed "evolutionarily plausible" groups [14]). Such risk framing seemed to lend concreteness to a risk estimate that may be absent from percentages or probabilities. Grounding in real-world groups enabled participants to visualize the reference class group (i.e., denominator group), and feel included within it—perhaps adding a personal connection to the statistic.

3.2.1.2. Theme 2: valuing the connection to medical research: Some participants stated that hearing quantifications of risk assured them that research has been conducted, providing a sense of authority beyond hearing verbal labels. This invocation of the broader community of medical researchers or scientists may confer credibility upon the source.

3.2.1.3. Theme 3: valuing transparent enumeration of outcomes: Many participants said they benefited from receiving –and thought they had a right to receive – statistical risk information, generally within an enumeration of all possible outcomes. This theme is congruent with the idea of patient-centered care.

3.2.2. Themes: obstacles to understanding numeric information—Our three remaining themes described obstacles to understanding numeric risk information. See quotations in Table 5.

3.2.2.1. Theme 1: lack of gradations in perception of uncertainty based on numbers: Some participants confided that they tend not to translate risk numbers into internal gradations of risk. These women seemed to perceive only three probability categories: certain yes, certain no, and uncertain outcome.

3.2.2.2. Theme 2: excessive reliance on numbers by clinicians in communicating risk: One concern voiced was that clinicians present numbers without sufficient explanation. Another was that when clinicians discuss numeric information with patients, it is at the expense of a genuine interpersonal connection.

3.2.2.3. Theme 3: confusion about related statistics: Some participants confused breast cancer incidence risk with the mammography outcomes risk, as the following illustrates:

> It said 1 in 70 . . . women get breast cancer and then at the end it asks you a question of 1 in a hundred? Um, I'm sorry. My brain is not computing what I should say now. (Participant #02)

Knowing this participant could only be referring to the risk scenario question discussed in Section 2.3.1 (as no questions were posed about the incidence statistics), her statement indicates she was trying to calculate the PPV of mammography from the cancer incidence statistics rather than from the mammography outcomes data. Her confusion may have been due to misinterpretation of the question, or to incomplete comprehension or forgetting of the flowchart data. After reviewing individual responses to the risk scenario question across the larger subject pool, we suspect other participants were similarly confused.

Participants who tried to use the breast cancer incidence statistic, approximately 1.4% (1 in 70), to answer the risk scenario question would likely have rounded down to "1" in the risk scenario question. Indeed, 18 of 69 participants answered "1" to this question—nearly as many as those who correctly answered "2" or "3" [22]. Two others typed "1 in 70", which we converted to a decimal and included in analysis. This further supports that some participants confused the breast cancer incidence statistics with those derived from the mammography outcomes graphic.

3.2.3. Risk graphics: recall of statistics and participant impressions—

Participants were asked to discuss any numbers they recalled from Mammopad. If none were recalled, they were probed about breast cancer incidence for average-risk women in their forties. Twelve participants were interviewed immediately after using Mammopad; the delay between presentation and recall was 15–30 min. The remaining nine participants returned for the interview after 14 to 122 days (median = 114, mean = 90).

3.2.3.1. Cancer incidence pictographs: Participants interviewed immediately after using Mammopad usually recalled risk information correctly from the cancer incidence pictographs (shown in Fig. 1), whereas those interviewed later tended to recall nothing or incorrectly recall a number at least 10% away from the correct answer. The results are cross-tabulated in Table 6. Ultimately, we observed little evidence participants recalled even the correct order of magnitude for breast cancer incidence rates from Mammopad after a delay of several weeks.

Participants liked the pink color in the cancer incidence pictographs and called the graphics simple and straightforward. They especially liked the use of woman-shaped icons:

> I remember really liking . . . the little pictures . . . The little visual 1 in 70. I think that's why I like it as an app as opposed to having a discussion or just reading an article. Somehow it helps to have visuals. (Participant #19)

Additionally, participants valued visual grounding in real-world groups, as described in Section 3.2.1.

3.2.3.2. Mammography outcomes flowchart: Only three participants voluntarily recalled information from the mammography outcomes flowchart: one recalled the figure nearly verbatim, another recalled only the denominator [1000], and one recalled that "just a very small amount" of abnormal tests were cancer.

Participants responded more variably to the flowchart than to the pictographs. Several participants said they had not read the entire flowchart during initial Mammopad use. The following quote from a participant who had had a previous abnormal mammogram illustrates one way this occurred:

> I didn't even read this part. The normal results, one cancer was missed, I never saw that one. 'Cause I focused on what my experience was. (Participant #05)

The participant followed one branch of the flowchart, and never returned to the top of the flowchart to read the remaining branch. Another participant described a different process

leading to incomplete reading of the depth of the flowchart (and failing to notice the missed cancer case):

> I think you mainly see the big [circles] . . . with the color. But you don't really pay attention to the ones down below." (Participant #10)

A third participant skipped the entire flowchart because the previous screen used slightly overlapping verbiage:

> I really didn't pay attention to the pictures because it, it said the same thing as what the previous page . . . So I kind of skipped the picture." (Participant #08)

Most participants said the information in the flowchart was valuable and important. No participants indicated unequivocal dislike for the screenshot, but those with mixed feelings either disliked the stick figure woman (including one who said "it reminds me a little bit of hangman." [Participant #04]), or thought the graphic took too long to understand.

When asked to state the main point of the flowchart, the following response was typical:

> this helps women have a realistic understanding of how the tests could come back, and then if the woman's test came back with a result that was out of the norm to not jump to the conclusion that it is cancer. (Participant #12)

Less commonly, participants thought the point was to reassure women there was only a small chance of having cancer at any given time.

Individual participants misinterpreted this graphic in unintended, potentially harmful ways. Because the flowchart node for normal test results was larger than the one for abnormal results, several thought the Mammopad designers aimed to focus attention on positive information and downplay negative information. Actually, we intended the size difference to reflect the lower frequency of abnormal results relative to normal ones. Although one participant supported our (alleged) decision to highlight positive information, another thought the abnormal branch should be "more pronounced so people are paying a little bit more attention . . . because the whole purpose is to detect it [cancer] early." (Participant #17). This participant also inappropriately inferred that people could rule out cancer by receiving a normal mammogram result.

4. Discussion and conclusion

4.1. Discussion

After using Mammopad, participants decreased their estimates of the PPV of screening mammography (as measured by a risk scenario question). Although the posttest estimate was much closer to the true value, this shift was probably partly attributable to confusion about different, related statistics (incidence of breast cancer over ten years vs. outcome of mammography on a specific day). We suggest two alterations that may mitigate this tendency for participants to confuse distinct (but topically-related) statistics: (1) explicitly contrasting differences between reference classes in the mammography outcomes flowchart and the breast cancer incidence pictograph, and (2) including comprehension checks after each Mammopad module. Further research is needed to determine if such changes would

effectively reduce confusion. Comprehension checks also would provide opportunity to communicate numeric information to participants answering incorrectly. Research on individual differences in statistical comprehension shows that, although strongly correlated with health numeracy, level of health-related graph literacy independently predicts understanding of graphical numeric presentations [14,37]. While graphic presentations are most helpful to low-numeracy people with high graph literacy, people with low numeracy scores and low graph literacy benefit most from non-graphical numeric presentations [37]. If users with low graph literacy fail a comprehension check of a quantity presented graphically, an alternate, numeric-only presentation might facilitate comprehension.

Through our content analysis, we discovered two categories of themes to characterize participants' experience with numeric statistical information, one describing reasons they find it valuable, and the other including obstacles to understanding. We advise PtDA creators to consider these reasons why participants value numeric and graphical formats in guiding design, to complement understanding gleaned from research comparing risk comprehension as a function of presentation format. Our results serve as a reminder that even when statistical information is effectively communicated, participants may not even produce estimates in the same order of magnitude after several weeks. This is concerning because many health decisions, such as screening decisions, have extended duration. Our thematic analysis suggests some value of being presented with numbers comes from trust being conferred upon the source. With this trust in place, participants may remember their decisions, even after forgetting risk statistics. Further, repeating Mammopad administration to a woman during her forties may mitigate effects of forgetting.

We discovered obstacles to understanding statistical information, including insensitivity to probability gradations. The decision we studied was affect-laden for participants. Research has shown that the more emotionally entangled the context of a decision was, the less sensitive participants were to variations in statistics [38]; however, most research on risk presentation is limited by use of fictitious scenarios. Another theme, worry that doctors focus on numbers without further explanation, highlights how important it is that statistical information is verbally explained and contextualized for participants by their clinicians, or within PtDAs.

Participants generally liked Mammopad's risk graphics. The simplicity of the breast cancer incidence pictographs and accompanying text, and their tendency to induce visualization of real-world groups of women (aided by use of woman-shaped icons) contributed to the graphics' appeal. The mammography outcomes flowchart graphic was more obtuse, and fewer participants fully comprehended it—and some said they disliked it for being too difficult to understand. We learned that some only read one dimension of the flowchart (e.g., down but not across). As such, we suggest that by adding an option to view an animated flowchart, Mammopad could guide participants through a complete reading, and in turn, increase comprehension. We are unaware of research on animated flowcharts, so this remains an untested hypothesis; however, effects of animation in other aspects of risk communication are complex [39–41].

This study's strengths include use of mixed methods and use of patients with real screening decisions using Mammopad in a clinic, as they might before a provider appointment when making the final decision about when to begin screening mammograms. Nonetheless, this study has several limitations. The recall portions would have benefited from a larger sample and targeted recruitment that ensured similar numbers of women with and without prior mammograms participated early and late. Our participants were primarily white, non-Hispanic rural-dwelling Oregonian women; caution should be taken in extending the results to other groups. Because lower education is associated with reduced health numeracy [22,42], it would be valuable for future work to oversample minimally educated participants. Ideally, qualitative data collection should continue until reaching thematic saturation. We halted recruitment for pragmatic reasons outside the control of this study. However, toward the end of analysis, we identified no new themes, suggesting we reached saturation.

The quantitative portion of this study used a pretest–posttest design, thus we cannot rule out the possibility that repeated testing caused the decreased estimates of PPV. Because we studied screening mammography decisions, extendibility to other decisions is limited. Although Mammopad presents harms of additional testing and treatment, these were not quantified, and thus fell outside scope of this study. Future evaluations of Mammopad should determine whether it effectively communicates such risks.

4.2. Conclusion

Our results indicated the presence of room for improvement in Mammopad's presentation of risk information; further research could determine whether the suggestions posed, such as comprehension checks and animation, would effectively bolster comprehension of complicated risk graphics. Even if risk information will be forgotten, many patients perceive benefit in receiving such information, particularly in pictograph format. Obstacles to understanding risk information, such as confusion between statistics, are important to address and mitigate in PtDA design.

4.3. Practice implications

Average-risk women deciding when to start regular breast cancer screening may benefit from using Mammopad before their clinic visit. Simple pictographs accompanied by text are easily understood and therefore may enhance shared decision-making discussions. When choosing visual aids to communicate statistical information, PtDA designers and providers should be aware of benefits and limitations of graphical representations—especially with more complex representations such as flowcharts. Incorporating comprehension checks into PtDAs would help identify misapprehension of graphically presented data and correct misunderstandings.

Acknowledgments

Appendix A

Interview guide

1. Opening/icebreaker: Before we discuss the mammography decision aid, Id love to hear more about you. Can you tell me a bit about yourself, such as family, job, education? Are you from the area originally?

2. Breast cancer knowledge/background: Could you tell me about what experience you had with people who got breast cancer? What had you heard about breast cancer before using the aid?

3. Health/prevention orientation: What would you say determines whether or not people stay healthy? Would you say that people's state of health is mostly within their control or mostly out of their control? *If needed*: Can you give me an example of something that is out of your control? In your control?

4. Impressions about Decision Aid Experience: How would you describe the decision aid to a friend who has heard a woman her age might need to get mammograms? How could the decision aid help a woman your age decide whether or not to have a mammogram? Sometimes you have an inner voice or dialog while youre doing things. Was this dialog saying anything while you used the decision aid?

5. Retention: What information or activities from the decision aid made the biggest impression on you? Can you tell me about anything that was hard to understand? What had value? Can you tell me about any numbers or statistics in the decision aid that were either helpful or difficult to follow? Prompt: What would you say is the biggest risk factor of breast cancer?

6. Numbers and risk: When a healthcare provider uses numbers in talking to you about your health, how does it make you feel? What thoughts come to mind? Why do doctors or nurses use numbers to talk about risk? Can you tell me about a time when information about risk or probability helped you make a health-related decision? (If never) What kind of quantitative information would be helpful to hear, if any?

7. Screen shots: I will show you some screen shots from the decision aid. For each screen shot, ask:

    • What thoughts do you have when you look at this screen shot? Can you describe what you see?

    • Do you like this screen shot? Why or why not?

- What do you think was the point of including this screen shot in the decision aid?

- How would you rate the importance of the information?

If time allows: If you were going to guess, why is there a 10 year difference in the guidelines for the US Preventive Services Task Force vs. the American Cancer Association?

Closing: Is there anything else you could tell me about, perhaps that I neglected to ask about?

Appendix B

Berlin Numeracy Test

1. Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips? _____ times out of 1,000.

2. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (odd = 1, 3 or 5)? _____ out of 50 throws.

3. In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS? _____ person(s) out of 1,000.

4. In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car? _____ %.

Appendix C

Appendix C

Table of individual characteristics of interview participants.

Participant ID	Age	Previous Mammogram	Berlin Numeracy Score (4=highest)
01	41	Yes	1
02	42	Yes	2
03	43	Yes	2
04	44	Yes	2
05	43	Yes	2
06	44	Yes	1
07	46	Yes	0
08	47[a]	Yes	3
09	43	Yes	2
10	48	Yes	2
11	47	Yes	2

Participant ID	Age	Previous Mammogram	Berlin Numeracy Score (4=highest)
12	43	Yes	2
13	49	Yes	2
14	45	Yes	4
15	40	No	4
16	42	No	Not administered
17	43	No	1
18	44	No	0
19	40	No	3
20	44	No	0
21	40	No	1

[a]Because only a birth year (and not a full birthdate) was available for Participant #08, her age was estimated.

References

1. O'Connor AM, Légaré F, Stacey D. Risk communication in practice: the contribution of decision aids. BMJ. Sep 27; 2003 327(7417):736–740. [PubMed: 14512487]

2. Volk, R.; Llewellyn-Thomas, H. Volk, R.; Llewellyn-Thomas, H., editors. The IPDAS background document: an introduction. Update of the International Patient Decision Aids Standards (IPDAS) Collaboration's Background Document [Internet]. 2012. Introduction http://ipdas.ohri.ca/resources.html

3. Barratt A, Trevena L, Davey HM, McCaffery K. Use of decision aids to support informed choices about screening. BMJ. Aug 28; 2004 329(7464):507–510. [PubMed: 15331483]

4. Stacey, D.; Légaré, F.; Col, NF.; Bennett, CL.; Barry, MJ.; Eden, KB., et al. Decision aids for people facing health treatment or screening decisions. In: The Cochrane Collaboration. , editor. Cochrane Database of Systematic Reviews [Internet]. John Wiley & Sons, Ltd; Chichester, UK: 2014. http://doi.wiley.com/10.1002/14651858.CD001431.pub4 [24.02.15]

5. Eden KB, Scariati P, Klein K, Watson L, Remiker M, Hribar M, et al. mammography decision aid reduces decisional conflict for women in their forties considering screening. J. Womens Health. Sep 11; 2015 24(12):1013–1020.

6. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. J. Natl Cancer Inst. Monogr. 1997; (22):63–67. [PubMed: 9709278]

7. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet. Dec; 2006 368(9552):2053–2060. [PubMed: 17161727]

8. Narod S, Sun P, Wall C, Baines C, Miller AB. Impact of screening mammography on mortality from breast cancer before age 60 in women 40 to 49 years of age. Curr. Oncol. Jul 3.2014 21(5):217. [PubMed: 25302030]

9. United States Preventive Services Task Force. Screening for breast cancer: U. S. Preventive Services Task Force recommendation statement. Ann Intern Med. Nov 17; 2009 151(10):716–726. [PubMed: 19920272]

10. O'Donnell J. Help me in my confusion: should we think more about mammography and colonoscopy as preference sensitive care? J. Cancer Educ. Dec; 2010 25(4):471–472. [PubMed: 21061111]

11. Woloshin S, Schwartz LM. The benefits and harms of mammography screening: understanding the trade-offs. JAMA. Jan 13.2010 303(2):164. [PubMed: 20068211]

12. Lipkus IM. Numeric, Verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med. Decis. Making. Oct; 2007 27(5):696–713. [PubMed: 17873259]

13. Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Appl. Cogn. Psychol. Dec; 1997 11(6):495–506.

14. Garcia-Retamero R, Galesic M. Using plausible group sizes to communicate information about medical risks. Patient Educ. Couns. Aug; 2011 84(2):245–250. [PubMed: 20728298]

15. Siegrist M. Communicating low risk magnitudes: incidence rates expressed as frequency versus rates expressed as probability. Risk Anal. Aug; 1997 17(4):507–510.

16. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ. Sep 27; 2003 327(7417):741–744. [PubMed: 14512488]

17. Gigerenzer, G. Calculated Risks: How to Know When Numbers Deceive You. Simon & Schuster; New York: 2002. p. 310

18. Cuite CL, Weinstein ND, Emmons K, Colditz GA. Test of numeric formats for communicating risk probabilities. Med. Decis. Making. Feb 12; 2008 28(3):377–384. [PubMed: 18480036]

19. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy in fluences risk comprehension and medical decision making. Psychol Bull. 2009; 135(6):943–973. [PubMed: 19883143]

20. Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. Health Aff. (Millwood). May 1; 2007 26(3): 741–748. [PubMed: 17485752]

21. Donelle L, Arocha JF, Hoffman-Goetz L. Health literacy and numeracy: key factors in cancer risk comprehension. Chronic Dis. Can. 2008; 29(1):1–8. [PubMed: 19036218]

22. Galesic M, Garcia-Retamero R. Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. Arch. Int. Med. Mar 8; 2010 170(5):462–468. [PubMed: 20212183]

23. Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. Health Psychol. 2009; 28(2):210–216. [PubMed: 19290713]

24. Schapira MM, Nattinger AB, McAuliffe TL. The influence of graphic format on breast cancer risk communication. J. Health Commun. Jul; 2006 11(6):569–582. [PubMed: 16950729]

25. McCaffery KJ, Dixon A, Hayen A, Jansen J, Smith S, Simpson JM. The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. Med. Decis. Making. Nov 10; 2011 32(4):532–544. [PubMed: 22074912]

26. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Educ. Couns. Dec; 2008 73(3):448–455. [PubMed: 18755566]

27. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. Cancer. Dec 15; 2008 113(12):3382–3390. [PubMed: 19012353]

28. Leung J, McKenzie S, Martin J, McLaughlin D. Effect of rurality on screening for breast cancer: a systematic review and meta-analysis comparing mammography. Rural Remote Health. 2014; 14(2):2730. [PubMed: 24953122]

29. Bellcross C. Further development and evaluation of a breast/ovarian cancer genetics referral screening tool. Genet Med. Apr.2010 12(4):240. [PubMed: 20395744]

30. Bellcross CA, Lemke AA, Pape LS, Tess AL, Meisner LT. Evaluation of a breast/ovarian cancer genetics referral screening tool in a mammography population. Genet Med. Nov; 2009 11(11): 783–789. [PubMed: 19752737]

31. [26.11.14] Georgia Breast Cancer Genomic Consortium. Breast cancer genetics referral screening tool (B-RST) [Internet]. Available from: www.breastcancergenescreen.org.

32. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med. Inform. Decis. Making. 2013; 13(Suppl. 2):S7.

33. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: ihe Berlin Numeracy Test. Judgm. Decis. Making. 2012; 7(1):25–47.

34. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann. Int. Med. Dec 1; 1997 127(11):966–972. [PubMed: 9412301]

35. Use the Berlin Numeracy Test [Internet]. RiskLiteracy.org; Available from: http://www.riskliteracy.org/researchers/ [16.05.13]

36. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. Qual. Health Res. Nov 1; 2005 9(9):1277–1288. [PubMed: 16204405]

37. Gaissmaier W, Wegwarth O, Skopec D, Müller A-S, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc. May; 2012 31(3):286–296.

38. Rottenstreich Y, Hsee CK. Money, kisses, and electric shocks: on the affective psychology of risk. Psychol. Sci. May 1; 2001 12(3):185–190. [PubMed: 11437299]

39. Witteman HO, Fuhrel-Forbis A, Wijeysundera HC, Exe N, Dickson M, Holtzman L, et al. Animated randomness, avatars movement, and personalization in risk graphics. J. Med. Internet Res. Mar 18.2014 16(3):e80. [PubMed: 24642037]

40. Zikmund-Fisher BJ, Witteman HO, Fuhrel-Forbis A, Exe NL, Kahn VC, Dickson M. Animated graphics for comparing two risks: a cautionary tale. J. Med. Internet Res. Jul 25.2012 14(4):e106. [PubMed: 22832208]

41. Han PKJ, Klein WMP, Killam B, Lehman T, Massett H, Freedman AN. Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. Patient Educ. Couns. Jan; 2012 86(1):106–113. [PubMed: 21377311]

42. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med. Care Res. Rev. Apr; 2007 64(2):169–190. [PubMed: 17406019]

Box 1. Risk scenario question

Jane is a woman in her 40s who is at average risk of developing breast cancer sometime during her life. She decides to have a mammogram to screen for breast cancer. She gets a call from her doctor saying that the result of the mammogram was abnormal and that she needs to have more tests to determine if she has breast cancer.

On a scale of 0–100, what are the chances that Jane has breast cancer, where 0 means she does not have breast cancer and 100 means she does have breast cancer.
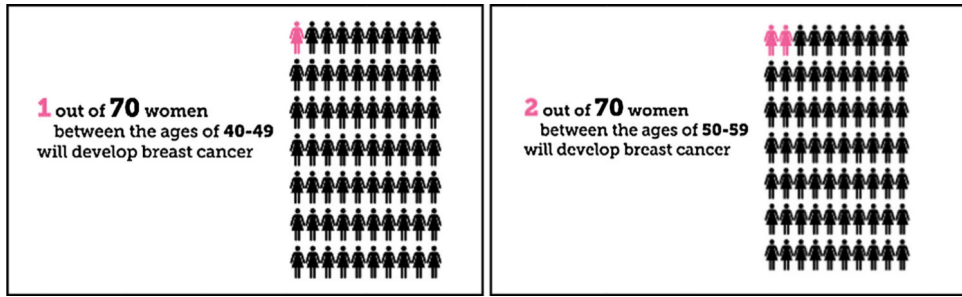
Fig. 1.
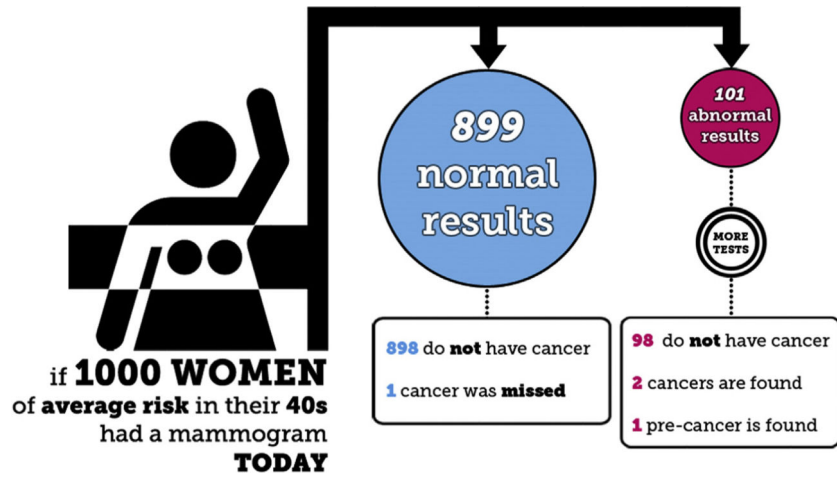Mammography incidence graphics from Mammopad.

Fig. 2.
Mammography outcomes flowchart graphic from Mammopad.

Table 1

Summary of participant characteristics.

	Quantitative phase	Interview phase
Total number	69	21
Age–mean (range)	45.0 [40–49]	44.3 [40–49]
Race[a]–N(%)		
American Indian or Alaska Native	2 (2.9)	2 (9.5)
Asian	3 (4.3)	1 (4.8)
Black or African American	0 [0]	0 [0]
Native Hawaiian or Pacific Islander	1 (1.4)	1 (4.8)
White	67 (97.1)	20 (95.2)
Declined to respond	1 (1.4)	0 [0]
Ethnicity		
Hispanic or Latino	3 (4.3)	2 (9.5)
Not Hispanic or Latino	66 (95.7)	19 (90.5)
Education level–N(%)		
Some high school	1 (1.4)	1 (4.8)
High school diploma or GED	18 (26.1)	3 (14.3)
Some college	18 (26.1)	9 (42.9)
Associate degree or technical certificate	16 (23.2)	3 (14.3)
Bachelor degree	12 (17.4)	4 (19.0)
Graduate degree	4 (5.8)	1 (4.8)
Health insurance		
Uninsured	10 (14.5)	1 (4.8)
Medicare/medicaid	15 (21.7)	6 (28.6)
Private insurance	44 (63.8)	14 (66.7)
Had at least one mammogram–N(%)	47 (68.1)	14 (66.7)

[a]Multiple responses were allowed for race.

Table 2

Summary of participant positive predictive value estimates before and after using Mammopad.

	Median (IQR)	Mean (SD)	Wilcoxon signed-rank test statistics[a]
Before	50 (28)	37.6 (16.1)	$Z(69) = 5.721$, $p < .001$
After	6 (21)	16.2 (20.2)	

Table 3

Comparision of participant positive predictive value estimates between experience grouping variables.

Variable	Measure	Status		Mann-Whitney U Test		
		Yes	**No**	**U**	**z-value**	**p-value** [*]
Previous Mammogram Status	Median score–before (IQR)	50.0 (25)	35.0 (30)	455	−.854	.393
	Median change score (IQR)	−20.0 (40)	−18.5 (39)	476	−.532	.594
Knowing a friend or family member with breast cancer	Median score–before (IQR)	40.0 (30)	50.0 (26)	525.5	−.391	.696
	Median change score (IQR)	−18.0 (38)	−24.0 (33)	490	−.862	.389

[*] The criteria for significance was set at $p < .05$, with Bonferroni correction. No tests reached significance.

Table 4

Value themes and quotations.

Quotation	Theme
"If you're thinking 1 out of 70, that's not very many, really. I guess for me that paints that picture of, if I'm at church or at a conference, you know, or smaller gatherings, one of the people in this room are likely to get breast cancer." (Participant #06)	Valuing grounding in real-world groups.
"Well, it puts it in a perspective that I can see. Like if you put it [the reference group] in under 100 numbers, it makes me feel like I'm part of that group." (Participant #17)	
"Numbers come from data, data comes from truth." (Participant #11)	Valuing the connection to medical research
"I think it's good [to hear numeric risk representations] because it shows that there's some kind of study maybe or some information that they're getting it from. It's not just like a thing they're coming up with themselves." (Participant #02)	
"I think that everything should be straightforward. I think that if there's a chance that something can go wrong that you should know what that something is and what the chance of it is." (Participant #16)	Valuing transparent enumeration of outcomes

Table 5

Obstacle themes and quotations.

Quotation	Theme
"I think if I heard them say 'you might get a headache', I would just take that. If they say, 'Oh, 25% of the time you get a headache', I'd be . . . It would mean nothing to me . . . . Just tell me that it could be a headache or it could not." (Participant #03)	Lack of gradations in perception of uncertainty based on numbers
"One in a thousand [may] not wake up. Again, I don't think that would have any meaning to me necessarily. I think I would have okay, well, there's a chance. That's what it would come down to, there's a chance." (Participant #05)	
"I generally want a better explanation than just a number. Like, explain what the number means. More in detail. Rather than what the odds are." (Participant #01)	Excessive reliance on numbers by clinicians in communicating risk
"I think it [the use of numbers]'s the language they speak . . . I don't think it's a language that I can make a connection with and relate to. I'm not offended by it, but I think that there could be a way that a doctor could communicate better." (Participant #12)	
"It said 1 in 70 . . . women get breast cancer and then at the end it asks you a question of 1 in a hundred? Um, I'm sorry. My brain is not computing what I should say now." (Participant #02)	Confusion about related statistics.

Table 6

Cancer incidence recall.

	Same-day recall[a]	Recalled on another day
Complete recall	11	0
Partial recall (missing reference class, or only remembered one of the two probabilities)	3	1
Recalled nothing	1	4
Incorrect probability recalled	0	4[b]

[a] One participant who would be in this category was not probed specifically about the incidence statistics and could not be analyzed.

[b] The four participants responded: 1 in 7, 1 in 4, 35%, and 30% for the incidence of breast cancer for women in their forties. The correct answer as presented in Mammopad was 1 in 70.