Evaluation of Risk Communication in a Mammography Patient Decision Aid

Capstone Project

By

Krystal A. Klein, Ph.D.

Presented to the Department of Medical Informatics & Clinical Epidemiology and the

Oregon Health & Science University

School of Medicine

in partial fulfillment of the requirements for the degree of

Master of Biomedical Informatics

March 9, 2015

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Krystal A. Klein, Ph.D.

"Evaluation of Risk Communication in a Mammography Patient Decision Aid"

Has been approved

Karen B. Eden, Ph.D., Associate Professor, Capstone Advisor

TABLE OF CONTENTS

Acknowledgements

Although my name appears alone as author of this capstone, this project could not have been completed without help. Karen Eden has been my advisor on this project and my Masters in Biomedical Informatics; her wisdom and support have been invaluable to me during the past two years. Joan Ash not only taught me qualitative research methods, but helped me refine this study's interview guide, provided helpful comments on this manuscript, and really served as an advisor to the project. I thank Dian Chase and Allison Sliter for serving as (unpaid) test participants in piloting my interview guide. I owe a debt of gratitude to Latha Kalaga and Lindsey Watson, who served as secondary coders and analysts for the interview phase of this study. I also thank and acknowledge the Oregon Rural Practice-Based Research Network, and especially its research coordinators, Vanessa Forro and Mark Remiker, who were critical in recruiting participants and maintaining good clinic relations.

It has been a long and somewhat circuitous road to where I am today. I owe thanks to my Ph.D. advisor, Richard Shiffrin, who recommended me for my first postdoc at Oregon Health & Science University, as well as to my former supervisor Misha Pavel, who gave me that position. I would like to thank Holly Jimison for encouraging me to apply for the National Library of Medicine fellowship that supported me in completing this project and my Masters in Biomedical Informatics, and for being my first advisor in Oregon Health & Science University's Department of Medical Informatics and Clinical Epidemiology and one of my best advocates.

Last and most, now and always, I must acknowledge the important contributions of my husband, Eric, who always believes in me and supports me.

Abstract

Objective. This study set out to use evaluate how well risk information was presented within a mobile patient decision aid app, Mammopad, and to provide a detailed characterization of benefits and shortfalls in quantitative risk communication scenarios, including the Mammopad graphics, where they occur.

Methods. Risk presentation quality was evaluated in three pieces: 1) A risk scenario question where participants answered a word problem with their quantitative estimate of the positive predictive value of screening mammograms for women in their forties (administered before and after using the decision app); 2) A thematic analysis of transcripts from interviews with a subsample of Mammopad participants about what they found valuable about quantitative risk information; and 3) A thematic qualitative analysis of transcripts from the same interviews about the interpretation of risk communication diagrams, including misperceptions/misinterpretations of data.

Results. Estimates of positive predictive value of mammography by participants (n=71) were more accurate after using the aid than they guessed before using it, although that was partly due to confusion between different risk statistics presented in the decision aid. Twenty-one participants enrolled in the interview phase: twelve participated in an interview session within an hour of using Mammopad, and nine others were recalled weeks or months after first using Mammopad. Negative themes concerning numeric risk presentation were: 1) lack of gradations in perception of uncertainty based on numbers, 2) numbers are sometimes provided as explanation, instead of a tool for explaining, 3) skepticism about the value of numeric information in light of forgetting, and 4) confusion about different statistics. Positive themes were: 1) valuing grounding in real-world groups, 2) valuing a connection to medical research, and 3) valuing transparent enumeration of outcomes.

Conclusions. The numerical risk graphics in Mammopad were well received and informative, although not memorable on a long-term basis. It may be helpful to administer Mammopad at multiple times during a patient's forties to refresh her memory. The more complicated of the two risk graphics in Mammopad was less often completely read or comprehended by participants, but cognitive tools such as animation could ease the cognitive burden of comprehending this information and improve understanding.

1. Introduction

A patient decision aid is an evidence-based tool designed to educate patients about their options concerning a specific medical decision--such as a choice between several treatments or screening options—as a supplement to, or preparation for clinical consultation. Because they focus on "preference-sensitive"[1] decisions—those where there is no "best" course of action across all patients—patient decision aids differ from general educational materials by helping patients understand how their own personal values relate to the features of the available decision options.[2] One of the touted benefits of decision aids is that that they allow more effective and balanced communication of risk information than occurs in typical clinical consultation.[3] Indeed, a 2014 Cochrane review found that across studies, patients who used patient decision aids in conjunction with typical care had superior knowledge and risk comprehension scores relative to patients who received care as usual.[4]

The present study partnered with an existing research project that evaluated changes in decision quality measures reported by patients after using Mammopad, a mobile device-optimized patient decision aid. Mammopad was developed to encourage women in their forties to understand and consider the costs and benefits of breast cancer screening options, to clarify their own values in relation to those options, and to empower them to discuss screening mammography with their healthcare providers. Recently, recommendations for routine mammography screening have been called into question in the face of equivocal evidence of benefit, *e.g.* from randomized trials investigating the impact of routine mammography on breast cancer mortality for average-risk women in their forties.[5–7] Indeed, while some organizations maintain these recommendations, the United States Preventive Services Task Force (USPSTF) recommendations to recommends against *routine* screening mammography for women in their forties, stating that the decision to begin biennial 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."[8] In other words, the USPSTF has found the decision to undertake biennial mammography as a preference-sensitive decision for women in their forties. This finding is supported by researchers with expertise in patient-centered care.[9,10]

The details of how to communicate risk to patients within a decision aid must be carefully considered on a case-by-case basis. There are many nontrivial decisions to make in creating a risk message. Currently, risk communication is as much an art as a science, despite a growing literature that maps the effects of risk formats [11–16], viewer characteristics[17–19], and chart types[20–22] on outcomes such as perception of risk levels, gist or exact recall of risk statistics, and subsequent decision outcomes (i.e., application of risk knowledge). This mixed-methods study aimed to record the impressions and reaction of rural patients, in their own words, to quantitative risk information, to gain understanding of the role and perceived importance that risk information presented in a decision aid has in the medical decision-making experience. Simultaneously, I aimed to investigate the effectiveness of Mammopad's specific risk communication graphics, by probing recall of key statistics and success at answering a word problem that required application of the presented risk information. I deliberately did not include any multiple choice tests, as is common in risk studies, because recognition memory is different from recall, and I considered recall more relevant to the goal of evaluating risk knowledge about the decision at hand.

2. Methods

I evaluated risk communication in Mammopad through triangulation using three approaches: 1) a quantitative experiment where participants answered a word problem about the positive predictive value of screening mammograms for women in their forties before and after using the decision app; 2) a thematic qualitative analysis of what participants found valuable about quantitative risk information in the app; and 3) a thematic qualitative analysis of interpretation of risk communication diagrams, including misperceptions/misinterpretations of data.

2.1. Participant Recruitment and Consent

2.1.1. Risk Scenario Participants

Participants in the Mammopad parent study, who were all women in their forties who were at average risk of breast cancer, were asked to answer the risk scenario question (described below in Section 2.3.1) immediately before and after using Mammopad. The parent study's participants were identified and recruited through chart review of three clinics identified through the Oregon Rural Practice-based Research Network (ORPRN) combined with telephone-based eligibility screening phone calls. They were screened to ensure that they were at average risk for breast cancer before they were permitted to proceed to using Mammopad. The precise details of recruitment and participant flow into the parent before-after study, including the risk screening process, were reported previously.[23]

2.1.2. Interview Participants

Early interview participants were a convenience subsample of Mammopad participants. After it was determined that all of these initial interviewees had previously had mammograms, the recruitment strategy was shifted to purposively recruiting Mammopad participants who had never had a mammogram. Because by this time the Mammopad study was also completing

enrollment, this meant recalling some women who had participated previously. All interview participants were recruited in the following manner: they were offered the opportunity to participate in a 30-40 minute interview evaluating the Mammopad app in exchange for a gift card.

2.1.3. Consent

Participants consented separately for the parent Mammopad study and the semistructured interview. Both consent forms and study protocols were approved by the Oregon Health & Science University Institutional Review Board in protocol IRB00007118.

2.2. Mammopad App

The Mammopad app included three main modules: an informational module on breast cancer, an informational module on mammography, and an interactive values assessment and clarification module.[23] Following completion of all modules of the app, a summary report was generated automatically and presented to participants; the report was then emailed to them if they provided a valid email address.

2.3. Data collection

2.3.1. Risk Scenario Phase

One risk scenario question was integrated seamlessly with the Mammopad decision aid to assess the participants' perception of the breast cancer risk associated with an abnormal mammogram result. This risk scenario question was posed to each participant immediately before using the decision aid. It read:

> 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 to 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.

The participant typed her response using the iPad's on-screen keypad. Immediately following the use of Mammopad, the question was again posed to the participant, and the response was collected in the same manner.

2.3.2. Interview Phase

An interview guide was developed in advance (*Appendix I*). Interviews lasted 30 to 40 minutes and were administered by me in private examination rooms at the ORPRN-affiliated clinics. The first part of the semi-structured interview focused on overall impressions of MammoPad and what aspects of the app participants found valuable or memorable. This segment of interview is being analyzed in a separate report. The second part of the interview probed for what statistics participants recalled from the app (focusing on the verbal explanations as well as memory for figures), and elicited participants' evaluations and explanations of what numerical information was or was not useful for them in the process of consuming health advice. For the final part of the interview, the participant reviewed screenshots from Mammopad and discussed the graphics and captions within those with the interviewer. Screenshots that explained quantitative concepts are analyzed in this manuscript; these included a pair of breast cancer incidence graphics for women in their forties and fifties, respectively (Figure 1) and a mammography outcomes graphic (Figure 2).

Figure 1: Mammography incidence graphics from Mammopad app.

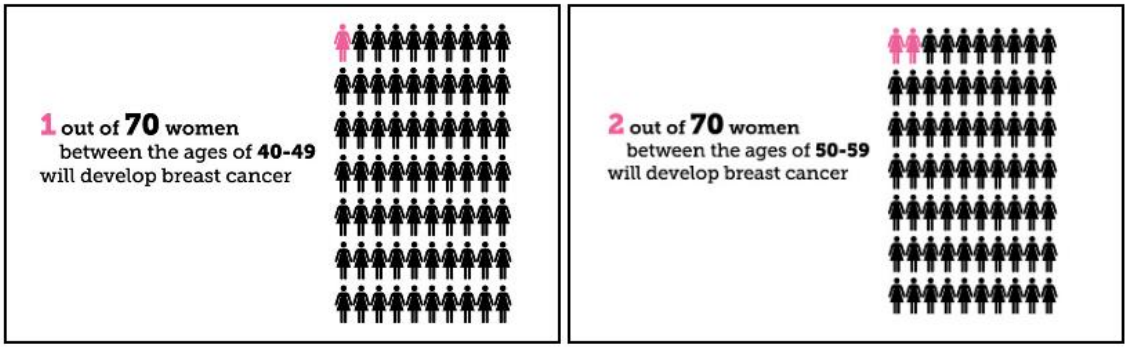
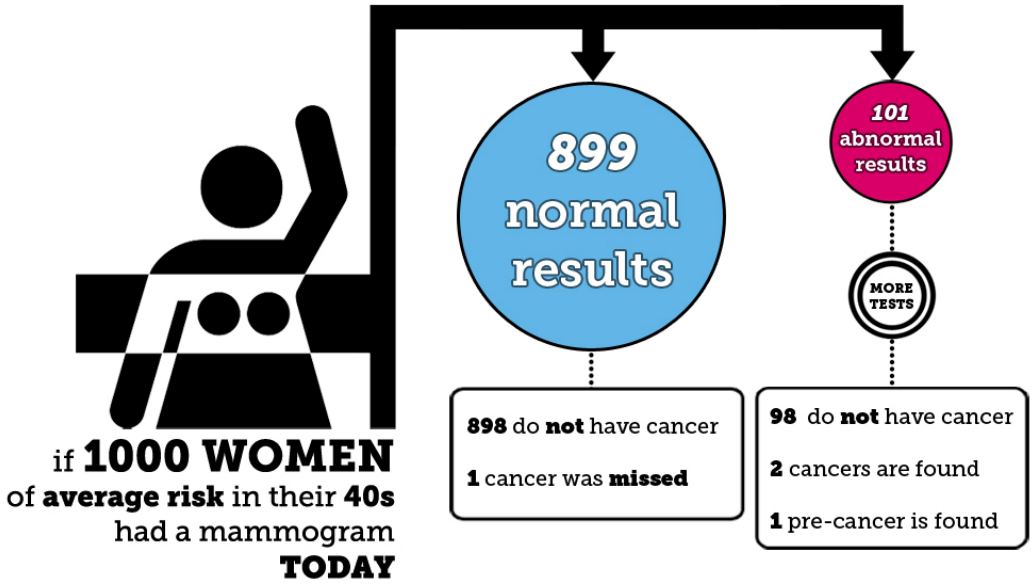


Figure 2: Mammography outcomes flow chart graphic from the Mammopad app.



After the interview, I asked participants to answer a short questionnaire evaluating their objective numeracy skill using the 3-minute Berlin Numeracy Test[24,25] (reproduced in *Appendix II)* recommended for participants with low numeracy[26]. The Berlin numeracy tests are

rigorously developed instruments that have demonstrated robust discriminability internationally and in diverse samples.[24] Importantly, score on the Berlin Numeracy Test has been found to more strongly predict comprehension of risk in everyday contexts than other available instruments. By measuring Berlin Numeracy score, I am able to provide descriptive data about our sample's command of quantitative concepts relevant to risk understanding and to confirm that people with different levels of numeracy were represented.

Interviews with all participants were audio-recorded, with verbal consent being obtained for recording. Interviews were transcribed by paid transcriptionists, and I 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). A Wilcoxon signed-rank significance test was planned to evaluate significance of change in risk estimates between pre- and post-Mammopad use.

2.4.2. Interview Phase

The semi-structured interviews were analyzed using a grounded theory approach. Analysis was completed using nVivo, version 10 for Mac (QSR International Pty Ltd, Doncaster Victoria, Australia).

2.4.2.1. Early Coding

One primary coder (myself) and two secondary coders contributed to the qualitative analysis. An initial review of three transcripts was undertaken by two coders using an open coding framework, which allowed analysts to add new codes at any time in the analysis. The coders then met to go over transcript coding and consensually converge on an initial coding scheme. The initial scheme included two levels, nodes and (optionally) subnodes. The coding

scheme, with the option for adding new codes, was used to code the remaining transcripts, with each transcript being coded by the primary coder and one secondary coder, and with review sessions resulting in consensus coding of each transcript recorded in nVivo software. These review sessions were held after every two to three transcripts coded. After each review session, the coding schema was updated and redistributed.

2.4.2.2. Refined coding

Following the consensus coding of 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 schema, the primary coder and one secondary coder reviewed all references that had been coded and drafted recommendations regarding changes to the schema. The analytic team then met to share their independent reports on each code/theme and to jointly connect and organize codes/themes, resulting in a final list of themes. Finally, quotes were selected to illustrate these themes.

3. Results

Demographic characteristics of the participants in the risk scenario phase and the interview phase are presented in Table 1.

Berlin numeracy score was assessed for the interview participants. One participant was not administered the Berlin Numeracy Test due to experimenter error. Possible scores ranged from 0 to 4 with no partial credit awarded. The sample mean score was 1.8 (sd=1.15).

Individual characteristics of participants in the interview phase are provided in *Appendix III*.

3.1. Risk scenario Phase

Non-numeric responses were converted to numeric responses when there was no ambiguity; for example, "Fifty-fifty" was converted to 50, and "twenty percent" was converted to 20. Six of the 75 participants responded to either or both the pretest and posttest with a non-numeric response that could not be directly translated to a single number and were excluded from further analysis. Thus, 69 response pairs were included in the experiment analysis.

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)
Minority – N (%)	5 (7.2)	0 (0)
Less than HS graduate/GED– N (%)	1 (1.4)	1 (4.8)
Had >= 1 mammogram – N (%)	47 (68.1)	14 (66.7)

Table 2: Summary of participant positive predictive value estimates before and after using the Mammopad app.

	Median (IQR)	Mean (SD)
Before	50 (28)	37.6 (16.1)
After	6 (21)	16.2 (20.2)

After using the decision aid, the participants reduced their ratings of how likely it was that Jane had cancer given an abnormal screening mammogram. This change was in the direction of a more accurate response. Based on the risk graphic presented in the decision aid, the target answer was 2.0 or 3.0, depending on whether or not the instance of "pre-cancer" was considered to be an instance of cancer. Before using Mammopad, participants gave a median response of 50, whereas the response after using Mammopad had a median of 6 (see Table 2). By a Wilcoxon signed-rank test, this was a statistically significant reduction; $Z(69) = 5.721$, $p < .001$; there were 49 negative shifts, 4 positive shifts, and 16 ties.

3.2. Interview Phase

Twenty-one of the Mammopad participants also participated in the interview. Enrollment for this portion of the study began toward the end of enrollment in the parent study. Early-enrolled participants in this study had their interview sessions immediately following the Mammopad session. Because the study team was short of the goal of enrolling 20 participants by the time the parent study concluded enrollment, later-enrolled interview participants returned for the interview weeks or even months after their initial sessions. Thus, I have gathered both immediate and delayed impressions of the decision aid.

3.2.1. Themes on numeric risk presentation

Several themes were expressed by participants about how they perceived the importance of quantitative risk information, both in Mammopad and generally. In describing them, I coarsely divide these into negative and positive themes.

Negative Theme 1: Lack of gradations in perception of uncertainty based on numbers

Some participants expressed a dislike of quantitative risk information in a manner that suggested they did not translate risk numbers into internal gradations of risk. For these women, uncertainty seemed unquantifiable, as illustrated by the following quotations.

> 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)

> 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)

Negative Theme 2: Numbers are provided as explanation, instead of a tool for explaining.

Several participants worried that numbers tended to be presented to them in healthcare situations without sufficient explanation, as the following quotation illustrates:

> 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)

Neglect to sufficiently explain numbers may reflect time pressures on behalf of clinicians. Alternately, it may reflect unawareness of how much difficulty some people have interpreting numeric risk information, as the following statement illustrates:

> I think it[the use of numbers]'s the language they speak… Sometimes 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)

Negative Theme 3: Skepticism about the value of numeric information in light of forgetting.

Two participants made metacognitive statements about how quickly they forgot or would forget specific numbers from the decision aid. For example, after misremembering the incidence of breast cancer for women in their forties as one in seven (an order of magnitude higher than reality) in an interview immediately after using Mammopad, one participant confided that upon hearing quantitative data about risks:

> I think I would quickly forget that number. (Participant #05)

Another participant, who returned for the interview weeks after using Mammopad, stated:

> I remember there being a lot of facts that I can't remember now, statistics which I thought were very helpful at the time, even though I do not retain anything with numbers. (Participant #19)

Negative Theme 4: Confusion about related statistics.

The interview transcripts contain evidence that some participants confused the breast cancer incidence risk data with the mammography outcomes risk data. The following quotation illustrates this confusion.

> 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 that the participant could only be referring to the risk scenario question discussed earlier in this paper (as no questions were posed about the incidence statistics), her statement indicates that she was trying to deduce the likelihood that our hypothetical patient, Jane, had cancer, from the cancer incidence statistics and not from the mammography outcomes data. To be sure, it is unclear whether this confusion was due to misinterpretation of the question itself, or to incomplete comprehension or forgetting of the data in the mammography outcomes flow chart. After manually examining the individual responses to the risk scenario question across the larger subject pool, however, I found potential evidence that other participants were similarly confused.

Mammopad presented breast cancer incidence for average-risk women in their forties as being 1 in 70. This is equivalent to 1.4%. Participants who tried to use the incidence rate to answer the risk scenario question would have rounded this down and given a response of 1 in the risk scenario question. Indeed, 18 participants did answer 1 to this question—nearly as many as those who gave a correct answer of 2 or 3 (22). Two others typed out "1 in 70", which we converted to a fraction and included in the analysis. This data provides additional evidence that participants had a tendency to confuse the breast cancer incidence risk with the risks derived from the mammography outcomes graphic.

Positive Theme 1: Valuing grounding in real-world groups.

A frequently-mentioned theme was that of grounding in real world groups. There is a concreteness about natural frequencies that does not exist in presenting abstract percentages or probabilities. One important element of the desire for grounding in real world groups seemed to be that it was important to be able to visualize the denominator group; for example, one participant said:

> 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)

Another participant expressed a similar sentiment, but added a point regarding the importance of feeling included within the denominator group, perhaps adding a personal connection to the statistic:

> 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)

Positive theme 2: Valuing the connection to medical research

For some participants, hearing quantifications of risk was valuable because it reassured them that research had been conducted, and this gave a sense of authority beyond hearing verbal labels, as illustrated by the following statement:

> Numbers come from data, data comes from truth. (Participant #11)

This may also have reassured participants that their clinician or clinic was trustworthy by virtue of having invoked the broader community of medical researchers or the scientific community, as is suggested by the next quotation.

> 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)

Positive theme 3: Valuing transparent enumeration of outcomes

Many participants considered themselves to gain benefit from receiving—or even thought they had a right to receive-- quantitative risk information, generally within an

enumeration of all possible outcomes. This theme is congruent with the idea of patient-centered care, and is illustrated below.

> 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)

3.2.2. Risk Graphics: Recall of Statistics and Participant Impressions

Participants were asked to discuss any numbers that they remembered reading in the Mammopad app. If they did not recall any specific numbers, they were probed as to what they thought the risk of breast cancer was for women in their forties. For twelve participants, the interview session occurred immediately after using the decision aid; although not explicitly timed, the duration of the delay between presentation and recall was estimated to be 15-30 minutes for these participants. The remaining nine participants returned to the clinic for the interview between 14 and 122 days (median = 114, mean = 90) following the initial use of Mammopad.

Cancer Incidence Pictographs

Participants who were interviewed immediately after using the decision aid tended to recall at least some of the quantitative information from the cancer incidence pictograph. Of the 11 participants in this category who were asked the question[i], seven met our criteria for perfect recall, in that they stated that the risk of breast cancer for women in their forties was 1 in 70 and that the risk for women in their fifties was 2 in 70 or double. Three participants recalled partial information, either that the incidence rate doubled between the forties and fifties or only the incidence rate for women in their forties. Finally, one participant could not recall any incidence rate from the app. Unsurprisingly, the nine participants who were interviewed between two

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

weeks and four months after using Mammopad displayed poorer recall of the incidence statistics. One participant recalled the frequency "1 in 70" but could not answer as to what the reference group was (i.e., women in their forties.) Four participants could not recall any numbers from the app. The remaining four participants recalled or guessed incorrectly; their four respective responses each exceeded the true incidence rate substantially: 1 in 7, 1 in 4, 35%, and 30%. From this small group, there is little evidence to support that participants recalled even the gist of the breast cancer incidence rates (1 in 70; about 1.4%) presented in the app after a delay of at least two weeks.

Participants responded positively to the cancer incidence pictographs (shown in Figure 1). Participants often brought up liking the pink color and finding the simple and straightforward. Another popular aspect of this graphic with participants was the use of icons shaped like women in the pictograph, as illustrated by the following statement:

> One thing I remember really liking about this app from before is 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)

Another participant's comments suggest what the appeal of the woman-shaped icons may have been for some viewers:

> I can see myself being a part of that group more than if it's not using a number at all, and maybe that's why it's helpful in that app to have the little people as part of, you know, the way you have the graphic there. (Participant #17)

The use of woman-shaped icons in the pictograph, therefore, may have contributed to women feeling that the information they were learning was grounded in real-world groups that they themselves could be part of.

Mammography Outcomes Flow Chart

Only three participants voluntarily recalled any information from the mammography outcomes flow chart: one recalled nearly the entire figure verbatim, one recalled only that there was a denominator of 1000, and one recalled that there was just "just a very small amount" of abnormal tests that turned out to be cancer.

When reviewing the graphic during the interview, participants responded more variably to the mammography outcomes flow chart than to the risk incidence pictographs. Several participants stated that they had not read the entire flow chart when going through the app. 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 seems to have followed one branch of the mammography outcomes flow chart, and never returned to the top of the flow chart to read the remaining branch. Another participant described a different process leading to incomplete reading of the flowchart diagram. After admitting that she had not noticed that one cancer was missed, she stated:

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

This quotation suggests that not reading through the complete depth of the flow chart is also a problem.

Another participant admitted to skipping over the entire flow chart because wording from the previous screen in Mammopad used some of the same 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)

It is true that there was some overlap in the content between the two pages, but most of the information in the flow chart was not provided in the previous page, which read only: "Mammography isn't perfect. There are tradeoffs. In fact, for every 1000 women like you that have a mammogram we expect…" (compare with Figure 2).

Most participants liked having the information provided in this graphic, and considered it to be important. Participants described liking the bold lines and color scheme, but there were mixed reviews of the stick-figure representation of a woman having a mammogram that had been integrated into the screenshot: some participants thought it was cute or clever, while others actively disliked it, and one thought "it reminds me a little bit of hangman."(Participant #04). None of the participants indicated unequivocal dislike for the screenshot, but those with mixed feelings either disliked the stick figure woman or thought that the graphic took too much time for them to understand.

When asked what the main point of the mammography outcomes flowchart was, the following quotation epitomizes how participants most typically answered:

> I think that 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 that to not jump to the conclusion that it is cancer. (Participant #12)

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

Individual participants also misinterpreted this graphic in unintended and potentially harmful ways. Four participants (#02, #14, #16, #17) thought the decision aid designers were trying to focus viewers' attention on positive information and downplay negative information, because the flow chart node for normal test results was much larger than the one for abnormal test results. In reality, this size difference was implemented in order to provide an additional visual clue reflecting the substantially lower frequency of abnormal test results relative to normal ones. [ii] Although one of these participants supported what she assumed was our decision to focus on the positive information, another participant thought that the abnormal branch of the flow chart needed more emphasis in order to ensure that it received due attention:

> I was gonna say the small pink one [node stating the frequency of abnormal results], I think you should have that a little bit more pronounced so people are paying a little bit more attention to that 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. The following exchange puts the extent of this misperception into context:

> PARTICIPANT #17: Just because you go in and get a mammogram doesn't mean you have cancer. It's just to show you that it's a good idea to rule it out.
>
> INTERVIEWER: So you can rule out cancer by having a mammogram?
>
> PARTICIPANT: Right.
>
> INTERVIEWER: Did you notice the missed cancer in here?

[ii] Although, as another participant pointed out, the actual areas that were ultimately used were misleading, from this perspective.

PARTICIPANT: Yeah. One out of a thousand.

This discussion suggests that in spite of having read and comprehended all the relevant frequency data about outcomes, the participant came to the false conclusion that mammography could rule out breast cancer.

4. Discussion and Conclusions

4.1. Discussion

This study aimed to characterize patients' comprehension, memory, and impressions of risk communication messages in a patient decision aid, Mammopad. After using Mammopad, participants tended to estimate that the positive predictive value of mammography (as measured by a risk scenario question) was significantly lower (median = 6%) than they guessed it to be before using the aid (median=50%). While the posttest estimate is much closer to the true value, this shift is probably partly attributable to confusion about different, related risks that were presented in Mammopad (incidence of breast cancer over 10 years vs. outcome of mammography on a specific day). Although unlikely to completely resolve confusion, I suggest two hypothetical alterations to the decision aid that may have mitigated this tendency for participants to distinct (but topically-related) risk enumerations: 1) including a discussion explicitly contrasting the differences between the reference classes in the mammography outcomes flow chart and the breast cancer incidence pictograph, or 2) including knowledge-test quiz questions after each of the breast cancer modules. Each of these has potential benefits and drawbacks. Further research would be needed to determine which, if either of these, would be most effective in reducing confusion.

Some of the themes that emerged from the qualitative interviews indicated a dubiousness or outright distrust of being told numerical risk estimates in medical contexts. The comments of some participants suggesting that they were insensitive to probability gradations is consistent with the results of a study by Rottenstreich and Hsee that found that the more emotionally entangled the context of a decision was, the less sensitive participants were to variations in statistics.[27] Indeed, medical situations and decisions are often quite affect-laden. Another theme, a desire for doctors not to focus on the numbers too much, is not surprising if

one subscribes to the prevalent, dual systems/processes models of decision making that suggest contributions by both a slow, logical component and a fast, affective or intuitive component.[28,29]

The risk scenario question results and recall results demonstrated that the Mammopad risk communication diagrams were reasonably effective in informing participants about true incidence of breast cancer and the outcomes of mammography, although the timeframe at which even gist memory of this information persists is disappointingly brief. This is of some concern because the type of medical decision being made, whether or not to be screened in one's forties, has an extended duration of up to a decade. Although some participants expressed doubt about the value of quantification of risk in light of propensity for forgetting, the thematic analysis suggests that some of the value of being presented with numbers is that it leads to trust being conferred upon the source. With this trust in place, participants may be more likely to at least remember their own decision on when to begin screening, even if they have forgotten all relevant risk statistics.

Participants generally liked the risk presentation graphics included in Mammopad. The breast cancer incidence pictographs were described by participants as easily comprehensible and visually appealing. The simplicity of these pictograph and the accompanying text, and the ease with which they could induce visualization of a real-world group of women (perhaps aided by the use of woman-shaped icons) seemed to contribute to the appeal of this particular set of graphics. The mammography outcomes flow chart graphic was more obtuse, and only some of the participants appeared to fully comprehend all the information it contained. Although there were some graphical design decisions that led to a subset of participants having mixed feelings about this graphic, other participants simply disliked the graphic because they deemed it too complicated or it took them too long to comprehend. In future versions of this decision aid or with similar risk communication graphics, the situation may be improved by further unpacking the information in this graphic into more of a narrative, as might be achieved through

animation. Doing so may improve the comprehensibility of the data to people with limited graph literacy, although this would likely come at the expense of forcing all participants to explore the chart in either a depth-first or breadth-first manner, to the exclusion of their own preferences.

A strong aspect of this study was that patients with real screening decisions used Mammopad in a clinic, just as they might before a provider appointment when making the final decision about when to begin screening mammograms. The study had some limitations, however. The recall portions of the study would have benefitted from a larger sample size so that a population estimate could be made for immediate and delayed recall of risk information presented in the decision aid. Because this study was tailored to mammography decisions, lessons learned from it may not transfer to contexts of treatment decisions or other screening decisions.

4.2. Conclusions

The numerical risk graphics in Mammopad were well received and informative, although not memorable on a long-term basis. It may be helpful to administer Mammopad at multiple times during a patient's forties to refresh her memory. The more complicated of the two risk graphics, the mammography outcome flow chart, was less often completely read or comprehended by participants, but further simplification seems imprudent considering how important the information provided in this graphic was considered by those who were able to comprehend it. Still, cognitive tools such as animation could ease the cognitive burden of comprehending this information and improve understanding.

References

1. O'Connor AM. Risk communication in practice: the contribution of decision aids. BMJ. 2003 Sep 27;327(7417):736–40.

2. Volk R, Llewellyn-Thomas H. The 2012 IPDAS Background Document: An Introduction. In: Volk R, Llewellyn-Thomas H, editors. 2012 Update of the International Patient Decision Aids Standards (IPDAS) Collaboration's Background Document [Internet]. p. Introduction. Available from: 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. 2004 Aug 28;329(7464):507–10.

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]. Chichester, UK: John Wiley & Sons, Ltd; 2014 [cited 2015 Feb 24]. Available from: http://doi.wiley.com/10.1002/14651858.CD001431.pub4

5. 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):63–7.

6. 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. The Lancet. 2006 Dec;368(9552):2053–60.

7. 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. 2014 Jul 3;21(5):217.

8. United States Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009 Nov 17;151(10):716–26.

9. O'Donnell J. Help Me in My Confusion: Should We Think More About Mammography and Colonoscopy as "Preference Sensitive Care'? J Cancer Educ. 2010 Dec;25(4):471–2.

10. Woloshin S, Schwartz LM. The Benefits and Harms of Mammography Screening: Understanding the Trade-offs. JAMA. 2010 Jan 13;303(2):164.

11. Lipkus IM. Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations. Med Decis Making. 2007 Oct;27(5):696–713.

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

13. Garcia-Retamero R, Galesic M. Using plausible group sizes to communicate information about medical risks. Patient Educ Couns. 2011 Aug;84(2):245–50.

14. Siegrist M. Communicating Low Risk Magnitudes: Incidence Rates Expressed as Frequency Versus Rates Expressed as Probability. Risk Anal. 1997 Aug;17(4):507–10.

15. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ. 2003 Sep 27;327(7417):741–4.

16. Gigerenzer G. Calculated risks: How to know when numbers deceive you. New York: Simon & Schuster; 2002. 310 p.

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

18. Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy Skill And The Communication, Comprehension, And Use Of Risk-Benefit Information. Health Aff (Millwood). 2007 May 1;26(3):741–8.

19. 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.

20. Schapira MM, Nattinger AB, McAuliffe TL. The Influence of Graphic Format on Breast Cancer Risk Communication. J Health Commun. 2006 Jul;11(6):569–82.

21. 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. 2011 Nov 10;32(4):532–44.

22. 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. 4th Int Conf Shar Decis Mak ISDM 2007 4th Int Conf Shar Decis Mak. 2008 Dec;73(3):448–55.

23. 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. Manuscript Submitted.

24. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring Risk Literacy: The Berlin Numeracy Test. Judgm Decis Mak. 2012;7(1):25–47.

25. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997 Dec 1;127(11):966–72.

26. Use the Berlin Numeracy Test [Internet]. RiskLiteracy.org. [cited 2013 May 16]. Available from: http://www.riskliteracy.org/researchers/

27. Rottenstreich Y, Hsee CK. Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk. Psychol Sci. 2001 May 1;12(3):185–90.

28. Tversky A, Kahneman D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychol Rev. 1983;90(4):293–315.

29. Sloman SA. The empirical case for two systems of reasoning. Psychol Bull. 1996;119(1):3–22.

Appendix I

Interview Guide

1. Opening/Icebreaker

Thank you so much for taking the time to be here today. I am doing this research to find ways to improve the health care experience for women and I could not do that without your help. With this research, I aim to evaluate the mammography decision aid. The questions I am going to be asking don't have any right or wrong answers; I am just interested in hearing from a variety of women who use the decision aid about their experience.

This is Krystal Klein and I am interviewing participant # _____ on _____

_____(date) at _____(time) at _____

_____(location). Can I confirm that you have read the consent form and signed it? And do I

have your permission to record this interview for research purposes only?

Before we discuss the mammography decision aid, I'd 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?

Optional: Does luck have an influence on your health? How about a higher power?

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 you're 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?

Can you tell me about radiation from mammograms?

6. <u>Numeracy</u>

When a healthcare provider uses numbers in talking to you about your health, how does it make you feel?

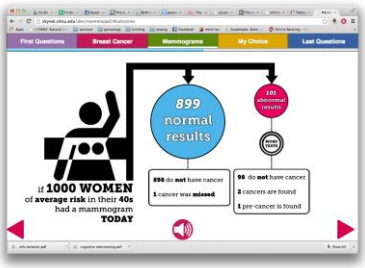
- Prompt: What thoughts come to mind?
- Prompt: Why do doctors or nurses use numbers to talk about risk?

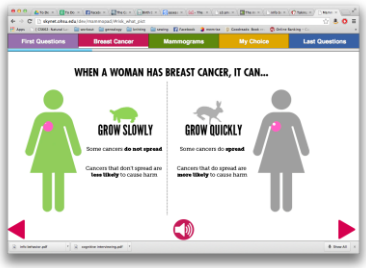
Can you tell me about a time when information about risk or probability helped you make a health-related decision?

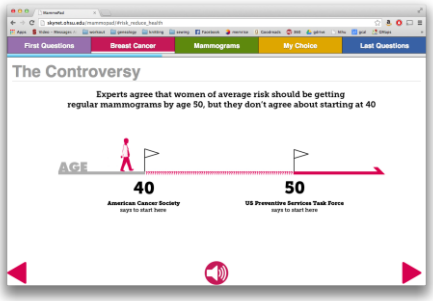
(If never) What kind of quantitative information would be helpful to hear, if any?

7. <u>Screen Shots:</u> I will show you some screen shots from the decision aid.

- 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 II

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 III

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*	yes	3
09	43	yes	2
10	48	yes	2
11	47	yes	2
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

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