European Journal of
Health Communication

ISSN 2673-5903
ejhc.org

Article

Communicating Risks with Interactive Visualisations and Reflective Tasks

DOI: 10.47368/ejhc.2025.102
2025, Vol. 6(1) 27-59

A Mixed-Methods Evaluation of a Mammography Screening Decision Aid

Jan Stellamanns
German Cancer Society (Deutsche Krebsgesellschaft e. V.), Germany

Keshav Dahal
School of Computing, Engineering and Physical Sciences, University of the West of Scotland (UWS), United Kingdom

Zita Schillmoeller
Department of Health Sciences, Hamburg University of Applied Sciences (HAW), Germany

Abstract
Static graphs of statistics are established visual aids in risk communication and decision support. Interactive information visualisations (InfoVis) and reflective tasks are supposed to enhance active processing, but the evidence is scarce and mixed. This mixed-methods research investigated the effectiveness and user experience of InfoVis and tasks in the context of mammography screening. In a web-based experiment prospective invitees of the screening program (*N* = 338; aged 30-49) tried a pre-tested web-based decision-aid with risk information either as text, static graph, or InfoVis with or without reflective tasks. The main outcomes were informed choice and risk knowledge, the latter operationalised according to the fuzzy-trace-theory. The accompanying qualitative evaluation with seven participants applied think-aloud protocols and focused interviews. There was no experimental evidence that InfoVis support risk knowledge or informed choice better than text or static graphs. There were even minor detrimental effects. The qualitative results showed problems with the InfoVis presenting risk of overdiagnosis, and negative reactions towards the- tasks. InfoVis processing was easy when the underlying concept was easy. While reflective tasks seem not advisable in this target group, limited and well-considered application of InfoVis with a low cognitive load can be an alternative, attention-directing visual aid format.

Corresponding author:
Jan Stellamanns: stellamanns@krebsgesellschaft.de

Keywords
Interactive visualisation, risk communication, mammography screening, medical decision-making, mixed methods.

A major challenge in health communication for lay information seekers, patients as well as health professionals is to get the numbers right. Difficulties understanding health statistics and risks have long been demonstrated and remain a prevailing concern (Jenny et al., 2018; Multmeier et al., 2014; Wegwarth et al., 2011). In a recent German survey, the interpretation of statistical information was the main issue for health professionals (Schaeffer et al., 2023). The importance of understanding quantitative information became obvious during the Covid-19 pandemic. Numerical epidemiological and medical concepts were omnipresent in the media with accompanying, sometimes interactive, dashboards and visualisations like "flatten the curve" graphs. The advantages and disadvantages of these formats have been critically discussed (Amidon et al., 2021; Ancker, 2020; Pérez-Montoro, 2022). While static risk graphs are established visual aids to support medical decision-making, recommendations regarding interactive formats are cautious due to scarce and mixed evidence (Bonner et al., 2021; Garcia-Retamero et al., 2012; Garcia-Retamero & Cokely, 2017; Stellamanns et al., 2017; Trevena et al., 2013, 2021). Information visualisations (InfoVis) – interactive, computer-supported visualisations of data – aim to support cognition, but they might even distort risk understanding (Card et al., 1999; Zikmund-Fisher et al., 2011).

Poor knowledge and misinterpretation of risk information can complicate the decision whether or not to participate in cancer screening, and adequate knowledge is a prerequisite to make an informed decision (Gigerenzer, 2014; Gigerenzer et al., 2009; Hersch et al., 2011; Marteau et al., 2001; Strech, 2014). Understanding these benefits and harms depends on numerical skills or numeracy: the ability to apply, process and grasp numbers. Higher numeracy is associated with favourable medical decision-making and more accurate risk comprehension in cancer screening (Koo et al., 2017; Malloy-Weir et al., 2016; Nelson et al., 2008; Reyna et al., 2009; L. M. Schwartz et al., 1997; Zikmund-Fisher, Mayman, et al., 2014). Some studies suggest that visual aids help to overcome comprehension problems of people with low numeracy skills (Fagerlin et al., 2011; Garcia-Retamero & Cokely, 2017; Garcia-Retamero & Galesic, 2010). In this context, graphical skills become important and may vary widely (Freedman & Shah, 2002; Keller & Siegrist, 2009; Shah, 1998, 2005; Zikmund-Fisher, Witteman, et al., 2014). Higher graph literacy is associated with more accurate risk estimates in studies evaluating static and interactive graphs (Gaissmaier et al., 2012; Nayak et al., 2016; Okan et al., 2012; Okan, Galesic, et al., 2015).

A better theoretical understanding of whether, why, when and for whom risk graphs and interactive formats help to communicate risks is necessary, and Padilla and colleagues proposed a cognitive dual-process model for decision-making with visualisations (Padilla et al., 2018; Stellamanns et al., 2017; Trevena et al., 2013, 2021). Following this, the fuzzy-trace theory suggests that quantitative information is rather processed through an arduous verbatim system, while risk graphs can facilitate processing by targeting the intuitive gist system (Reyna, 2008, 2012; Reyna & Brainerd, 2011). Another inconclusive interactive feature is reflective tasks: while experts discourage letting the reader do maths, active processing with numerical tasks have shown benefits in risk communication (Natter & Berry, 2005; Okan, Garcia-Retamero, et al., 2015; Trevena et al., 2021).

The Evidence on Interactive Visualisations and Reflective Tasks in Health Risk Communication

So far, two studies have not indicated advantages of InfoVis to communicate health risks to a lay audience; another one even showed detrimental effects. Ancker et al. (2011) assessed the effects of interactive icon-arrays on risk estimates, risk feelings, and decision-making compared to static graphs in a preventive, hypothetical scenario. The authors assumed deeper cognitive processing by interacting with the graphs. With no main statistical effects, they concluded that interactive graphs might be useful to shift interest to quantitative information (Ancker et al., 2011). Mason and colleagues (2014) tested interactive-to-static bar charts representing a hypothetical personal cardiac risk increase in web-based experiments. The recall of key risk information one week later was not statistically different compared to static graphs or numbers only. In a second qualitative study based on think-aloud protocols the authors reported a higher user involvement with interactive graphs (Mason et al., 2014). In another hypothetical scenario, Zikmund-Fisher and colleagues (2011) found that participants applying interactive icon-arrays were less likely to choose a favourable treatment option and that substantially more people dropped out than with static icon-arrays. The authors suggested a burdening and distracting effect due to interactive icon-arrays, considering these were not tested prior to the online experiment (Zikmund-Fisher et al., 2011). Reflective tasks are another means to enhance risk understanding. In a paper-and-pencil-based experiment, Natter and Berry (2005) showed that active processing can be encouraged, and more accurate risk estimates achieved through actual graph drawing and reflective questions. Okan and colleagues (2015) showed similar positive effects in hypothetical medical decision scenarios using icon-arrays and reflective tasks.

Some limitations need to be addressed. Firstly, recent research has stressed the importance of graph literacy for evaluating visual aids, which was not assessed in the first two studies (Ancker et al., 2011; Etnel et al., 2020; Mason et al., 2014; Trevena et al., 2021; Van Weert et al., 2021). Secondly, hypothetical decision-making scenarios prevailed, consistent with reviews on medical decision-making, where the authors challenged the transferability of the results (Blumenthal-Barby & Krieger, 2015; Patel et al., 2002). Thirdly, the research neglects theoretical approaches already mentioned (Lipkus, 2007; Lipkus & Hollands, 1999; Stellamanns et al., 2017; Trevena et al., 2013). While there is good evidence that static graphics are beneficial in communicating health risks, it is unclear whether and when additional interactivity helps or worsens cognitive processing (Garcia-Retamero et al., 2012; Garcia-Retamero & Cokely, 2017).

Objectives and Hypotheses

Our aim was to clarify whether InfoVis and reflective tasks can support lay people's understanding of risk. The primary objective was to assess the effectiveness of interactive icon-array and line graphs alone or in combination with reflective tasks in a quasi-realistic decision-making scenario with prospective users. The scenario was the decision to participate in mammography screening for women aged 30 to 49 years before they were invited to participate in the German national programme at the age of 50. The further objectives were to apply an appropriate processing theory, the fuzzy-trace-theory, and explore the user experience, information processing, and decision-making with InfoVis and reflective tasks. We assumed that the effectiveness of the interactive features can best be assessed with a quantitative approach, while the questions beyond that require an explorative qualitative approach. Therefore, we applied a mixed-methods evaluation design. For an online experiment we operationalised risk knowledge according to the fuzzy-trace theory. We hypothesised that:

> *H1: Risk knowledge increases with graphical and interactive features comparing risk information as text, statistic graphs, InfoVis without and with reflective tasks, while controlling for the effects of numeracy and graph literacy.*
>
> *H2: Effects on risk knowledge can be explained by differences in gist and verbatim processing.*
>
> *H3: Informed choice is enhanced by the application of InfoVis and reflective tasks.*

Qualitative methods aimed at exploring users processing of risk information and decision-making with InfoVis and reflective tasks.

Methods

For a wide application of InfoVis to inform lay audiences about health risks, there should be confidence that they do not distort messages and work equally well or better than other risk communication formats like text, tables, or static graphs. The safety and effectiveness of an intervention can most robustly be assessed with an experimental, randomised controlled research design, which is the gold standard in health research and a cornerstone in the evaluation of InfoVis (Alexander et al., 2015; Carpendale, 2008; Plaisant, 2004). Because it is unclear how and why InfoVis might support processing of health risk information and medical decision-making, we added explorative methods to ameliorate pure performance measures with context information, for example to explore the data exploration experience and cognitive processing (Carpendale, 2008; Isenberg et al., 2008; Plaisant, 2004; Riggin, 1997).

Mixed methods research (MMR) basically requires one qualitative and one quantitative research component, and allows to investigate confirmatory and exploratory research questions in one study (Creswell & Plano Clark, 2011; Greene et al., 1989; Schoonenboom & Johnson, 2017; Venkatesh et al., 2016). Our main research component or strand – according to MMR terminology – was the quantitative evaluation of InfoVis and reflective tasks in a randomised online experiment (Creswell & Plano Clark, 2011). Before we ensured comprehensibility and usability of the InfoVis, reflective tasks, and the experimental setting and quantitative measures in two qualitative pre-tests. In parallel to the online experiment, we qualitatively investigated the users experience, processing of information, and decision-making in a qualitative evaluation. We mixed methods to achieve complementarity and to investigate overlapping phenomena – risk knowledge and decision-making (Greene et al., 1989). The mixing occurred by combining and integrating the results from quantitative and qualitative data on the level of interpretation and conclusion to derive meta-inferences (Anthony & Joseph, 2014; Creswell & Plano Clark, 2011; Schoonenboom & Johnson, 2017; Venkatesh et al., 2016).

Participants, Recruitment, and Sample Sizes

We recruited female participants from 30 to 49 years old before they were invited to the German mammography-screening program. In the nationwide program, women aged 50 to 69 are invited every two years along with a printed decision aid (IQWIG, 2017; Malek & Kääb-Sanyal, 2016). We applied various recruitment strategies with non-probabilistic and convenience sampling methods, which were adapted to the qualitative and quantitative approaches. For the qualitative strands, the first author recruited participants from friends and acquaintances, directly or through snowball sampling. Eligible participants were contacted face-to-face, via phone or via email. The overall aim of the study was described and how the test would be conducted. Targeted participants were asked to further distribute the email invitation accompanied by an electronic information leaflet.

The determination of the sample size for the pre-test followed conventions of qualitative usability testing, assuming that about five participants can detect the most and most severe usability issues when the test is iterated, aims to detect major issues, and targets a distinct group with similar needs (Alroobaea & Mayhew, 2014; Nielsen & Landauer, 1993; Spool & Schroeder, 2001; Virzi, 1992). We conducted two qualitative usability pre-tests. Six women participated pairwise in the first test at the usability laboratory of the University of Applied Sciences Hamburg (HAW). For the second usability test, the main author visited seven participants at their homes alone ($n = 3$) or in pairs ($n = 4$) and conducted the test on the participants' private computers.

For the qualitative evaluation, the sample size was determined to obtain saturated, rich and adequate qualitative data (Morse, 1995; Williamson, 2018). The qualitative data should reflect varying degrees of information processing, understanding, attitudes, and decisions for and against mammography screening with InfoVis and reflective tasks. Because the qualitative evaluation was confined and focussed on the experimental stimuli and specific outcomes, we assumed that saturation can be achieved with five to ten interviews (Guest et al., 2006). The final sample consisted of seven single interviews.

The sample size for the experiment was based on similar studies with risk knowledge as the main outcome. A systematic review did not provide directly comparable studies to estimate an accurate sample size for the effect of interactive visualisations (Stellamanns et al., 2017). Hence, the effect was indirectly concluded from controlled studies evaluating static graphs, and decision aid studies with a mammography screening scenario and non-controlled design (Berens et al., 2014, 2015; Mathieu et al., 2010). A benefit of 10% in risk knowledge due to static graphs was concluded from a systematic review (Stellamanns et al., 2017). Applying this difference to data from the non-controlled mammography screening studies, a moderate effect size was calculated with the www.psychometrica.de website. We assumed that this effect size might be exaggerated for the comparison of textual risk information, static and interactive visualisations, and chose a small to moderate effect size ($f = 0.2$) for the main effect of our study. With the G*Power program we calculated a sample size of $n = 277$ for four independent groups for a main effect in the ANOVA, with an alpha error of .05 and a test power of 0.8.

Participants for the online experiment were recruited with diverse strategies to reduce the risk of selection bias, although without random selection. First, friends of the main author were asked to participate and distribute the invitation to the experimental website. Second, the study was advertised on a website of the German Cancer Society regarding mammography screening information. Third, the invitation to the study was placed on Facebook webpages of institutions which provide information about breast cancer such as the German Cancer Information Service, and self-help groups like Breast Cancer Germany. Fourth, we contacted eligible students of the HAW Hamburg via official student mailing lists. Further, we displayed information leaflets at the counselling office of the Hamburg Cancer Society. About half of the targeted sample size was achieved with these strategies, but further recruitment was slow and insufficient. Finally, a web-based research panel was commissioned to complete the sample.

Conditions and Design

The current decision aid of the German mammography screening program was the base for this evaluation and the InfoVis. It was designed to provide evidence-based, balanced, and transparent information about mammography screening. It came into effect at the end of the 2017 after a comprehensive qualitative and quantitative evaluation (Gemeinsamer Bundesausschuss, 2014, 2017; IQWIG, 2017). The decision aid accompanies the invitation letter that is sent out to eligible women aged 50 to 69 every two years (Das Mammographie Screening-Programm, n.d.).

The decision aid contains general information about breast cancer, the screening process, advantages and disadvantages of screening, alternatives, limitations, and privacy issues. Statistical information is presented in four places in various formats: (1) a table showing breast cancer incidence and mortality within the next ten years for women aged fifty, sixty and seventy, (2) a frequency tree depicting results of the screening for a thousand women including the numbers of negative, positive, true-positive, false-positive women, (3) an icon array graph with thousand dots illustrating the difference of breast cancer mortality with and without mammography screening, and (4) the same format showing the difference of breast cancer diagnosis with and without mammography screening to illustrate the number of overdiagnoses.

We transferred the decision aid to an experimental online platform. The outlined statistical information were the experimental stimuli that we manipulated. For the first experimental group (Text-only), we described the statistics with only text. The second experimental group (StaticVis) had the same formats as in the printed decision aid. For the third group (InfoVis), we transformed the statistical information into InfoVis with the website infogram.[1] The risk of breast cancer and breast cancer mortality for different age groups was designed as an interactive line graph. This format is most suitable to depict trends over time (Garcia-Retamero & Cokely, 2017). Moving over the data points in the line graph revealed the exact numbers in pop-up windows. We converted the other three points into interactive icon-array graphs with thousand female icons as denominator. The frequency tree with the screening results was converted into four subsequent and labelled icon-array slides. The difference of breast cancer mortality and breast cancer diagnosis with and without screening was depicted with two icon-array slides each. Participants could switch between the slides by mouse-clicking. Again, moving with the mouse over the icons revealed additional textual information in a pop-up window. In the non-interactive experimental groups this information was shown on extra webpages. Members of the fourth experimental (InfoVis+) group were additionally asked to solve reflective tasks after three of the four InfoVis: to calculate 1. the difference of breast cancer incidence between the age of 50 and 60 years, 2. the difference of breast cancer mortality with and without the screening (screening mortality benefit), and to calculate 3. the difference of breast cancer diagnosis with and without the screening. These numbers were provided in the non-interactive experimental groups.

In summary, there were four experimental conditions differing in the format of the statistical information: the first group with text only information, the second with static graphs and tables, the third with InfoVis, and the fourth with InfoVis and reflective tasks. In the qualitative pre-test and the main qualitative evaluation, we only tested the fourth group.

Randomisation and Masking

Participants of the online experiment were randomised after giving informed consent on the first webpage. To avoid biased response, participants were only disclosed about the broader aim of the study – to test and improve web-based mammography screening information. The randomisation into groups was concealed. In the qualitative strands interviewees were also masked about the specific aims of the study, although partly directed towards the InfoVis and tasks later in the focussed interview. The analyses were not masked, and all conducted by the first author.

Ethical Aspects

Before participating everyone was informed about the procedures and the general aim of the study. Each participant provided written informed consent or consented by clicking a button to agree to the described information and procedures. Participation in the online experiment was either completely anonymous without compensation or compensated and anonymised for members of the online panel due to the regulations of the provider. Participants in the qualitative pre-tests and evaluation were not compensated. Their identity was pseudonymised for the interview transcription. An ethics proposal for the mixed-methods study protocol was approved by the ethic committees of the University of the West of Scotland and the University of Applied Science Hamburg.

Data Collection and Analyses

For the qualitative pre-tests, we applied the think-aloud method. For the main qualitative evaluation, we additionally conducted focused interviews. In the web-based experiment, questionnaire data were collected via the online platform. The main outcome measures were risk knowledge and informed choice. We controlled for several covariates, mainly numeracy and graph literacy.

Pre-Test Data Collection and Analysis. The aim of the qualitative pre-tests was to detect any usability issues and comprehension problems regarding the InfoVis, the reflective tasks, and the questionnaires. The think-aloud method motivates participants to verbalise their current cognitive processes. It is a standard method in usability and visualisation research to gain rich and contextual information regarding problem-solving and knowledge acquisition (Carpendale, 2008; Isenberg et al., 2008; Someren et al., 1994). The qualitative data were based on transcripts of recorded verbal expressions and reactions of the participants towards the web-based decision aid. The main author transcribed the voice recordings with MAXQDA software (n.d.).

Qualitative Evaluation Data Collection and Analysis. For the main qualitative evaluation, we added focused interviews to the think-aloud method. The latter enables access to the working memory while executing a task, but it is less appropriate to investigate more intricate cognitive and mental processes of the long-term memory, attitudes, beliefs, emotions, and experiences, which play a crucial role in women's perspective on mammography screening (Hersch et al., 2016; Payne & Westerman, 2003). Focused interview is a qualitative method to explore and analyse the response to specific media stimuli and relies on retrospection. It can

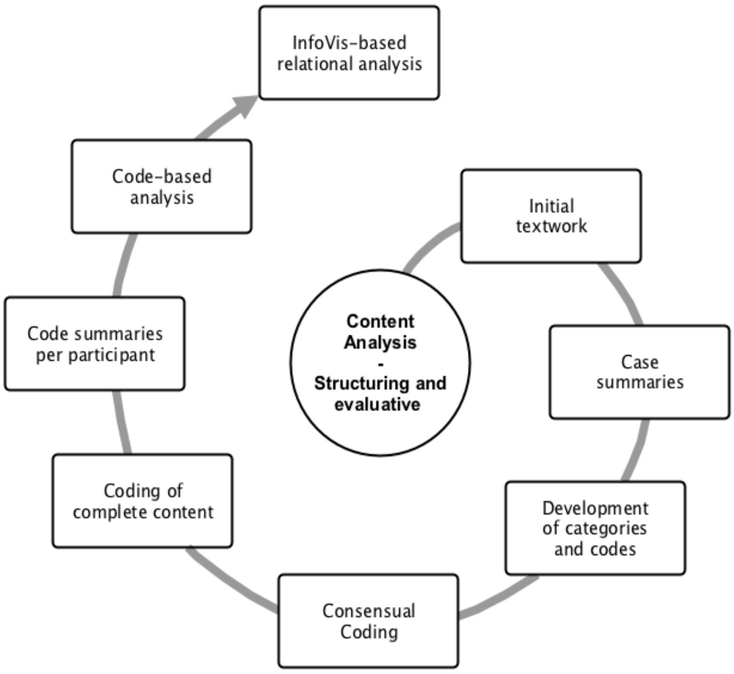

Figure 1. Process of the Structuring and Evaluative Content Analysis

compensate for the shortcomings of the think-aloud method and complement experiments to identify processes involved in experimental effects (Merton et al., 1990; Someren et al., 1994).

The interviews follow a semi-structured, guided manual. To allow for varied and unpredictable responses, a non-directive approach is essential. First, unstructured questions are prioritised, but structured questions can be integrated in the later stages. (Merton et al., 1990). The main author conducted the evaluative content analysis to develop codes and categories (Figure 1) (Kuckartz, 2018). After developing preliminary codes and categories, these were tested by consensual coding with another researcher in a sample of two interviews. Disagreement was discussed, the categories refined, and all interviews were coded accordingly. Further, we used relational analysis as an additional technique to relate codes and categories and present a relational model as a diagram and visual summary of the analysis (Busch et al., 2005; Robinson, 2011).

Quantitative Measures and Covariates for the Online Experiment. Understanding the benefits and risks is a prerequisite to making informed medical decisions, and knowledge is an essential component of informed choice (Elmore, 2016; Reyna, 2008). Therefore, our main outcome measures were risk knowledge and informed choice about mammography screening. Informed choice measures are mostly multidimensional including the dimensions knowledge, attitude, and actual or intended choice (Ames et al., 2015; Hersch et al., 2016). An informed choice is based on sufficient knowledge, combined with either a negative attitude and refusal, or a positive attitude and participation (Marteau et al., 2001; Mathieu et al., 2010; Michie et al., 2002). We adopted knowledge, attitude and intention items from studies investigating screening decisions, especially regarding mammography screening in Germany or that applied the fuzzy-trace theory (Berens et al., 2014, 2015; IQWIG, 2017; Reder et al., 2019; Reder & Kolip, 2015, 2017; Smith et al., 2010).

Our risk knowledge scale consisted of 12 multiple-choice items; all were related to quantitative screening information depicted in the InfoVis. Five gist items had verbatim counterparts, resulting in five pairs. Gist questions asked for a qualitative assessment, for example the likelihood of a breast cancer diagnosis with and without screening. Verbatim questions asked for an exact number, for example the difference of diagnoses with and without screening. Two knowledge items were neither gist nor verbatim (s. Appendix). The correct answers were summarised into a continuous scale from 0 to 12. For informed choice, we dichotomised the risk knowledge scale at the median into sufficient and insufficient knowledge according to prior studies (Berens et al., 2014, 2015; Marteau et al., 2001; Mathieu et al., 2010). Attitude comprised of four items on a 5-point Likert scale. Respondents were asked whether mammography screening is more or less important, comfortable, rather a good or a bad thing, and rather beneficial or disadvantageous. Values above the median on the sum score from 0 to 16 indicated a positive attitude. All participants with sufficient knowledge and either a positive attitude and intention to participate or a negative attitude and no intention were categorised as making an informed choice.

To compare the attractiveness and accessibility of the static graphs and InfoVis, we selected three items from a study by Gaissmaier and colleagues (2012). Participants were asked to assess whether the graphs were attractive or unattractive, comprehensible, or incomprehensible, and helpful or not helpful on a 5-point Likert scale.

To measure numeracy as a covariate, we followed the suggestion by Cokely and others (2012; 2014) and combined two instruments covering a wide range of numeracy skills, the Berlin Numeracy Test and the Numeracy Scale by Schwartz and colleagues (1997). Both scales combined show favourable distributional properties, while reliability was not reported and is generally low in various numeracy scales with Cronbach's alpha values below .70.

To assess the effect of graph literacy, we integrated it as a covariate. The Graph Literacy Scale (GLS) specifically measures health-related graph interpretation skills with acceptable reliability in the United States (n = 492) and Germany (n = 495) (Cronbach's alpha .79 resp. .74) (Galesic & Garcia-Retamero, 2011, 2013). The authors point out that the original 13-item scale might be too time-consuming, demanding and may provoke negative reactions (Garcia-Retamero et al., 2016). We adapted a 4-item GLS version by Okan and colleagues (2019) with comparable properties. It can predict better understanding with visual aids with an acceptable internal consistency, although low reliability (Cronbach's α = .53).

Previous knowledge or experience with mammography, screening or breast cancer can influence the decision-making in breast cancer screening. Expected harm, inconvenience or a recent breast cancer screening, lack of knowledge about breast cancer, and the influence of family and close friends can be predictors of non-participation in mammography screening (Aro et al., 1999; Kaltsa et al., 2013; Lagerlund, Hedin, et al., 2000; Lagerlund, Sparén, et al., 2000; Sterlingova & Lundén, 2018). To control for these effects, we asked participants whether they have ever undergone mammography, whether they have ever looked for information on mammography screening, or whether they knew a person who has or had breast cancer. Further participants were asked which device they used for the experiment, their date of birth, and educational background.

Statistical Methods. Our main quantitative research question was whether the InfoVis and the reflective tasks affect risk knowledge. We hypothesised that the interactive stimuli support risk comprehension and that the risk knowledge scores differ statistically significantly between the experimental groups. We applied an analysis of variance (ANOVA) to investigate the overall effect. We evaluated the prerequisites of ANOVA with the Levene's test for homogeneity of variance. Normality of the dependent variables was investigated visually with histograms, QQ-plot of the residuals of the model, and the Shapiro-Wilk test.

ANOVA only gives an overall effect but no effect sizes between the groups. Hence, we further applied planned contrast ANOVA with the following orthogonal contrasts: 1. Text-only vs. all three graphical groups, 2. StaticVis vs. InfoVis and InfoVis+ combined, and 3. InfoVis vs. InfoVis+. For each contrast we calculated the effect size *t* (Field et al., 2012). Since the planned contrasts cannot reflect all possible comparisons, we also conducted post-hoc methods with the Tukey and Dunnett test.

We included numeracy, graph literacy and other modifying variables, in stepwise analyses of covariance (ANCOVA), factorial ANCOVA and independent mixed models, when both categorical and interval co-variables were integrated. Independence of the covariate and the interventional effect, and the homogeneity of regression slopes were assessed according to Field et al. (2012). We applied the same planned contrast as for the ANOVA.

To ascertain whether risk knowledge rather relies on gist knowledge and thus reflects the fuzzy trace theory we repeated the risk knowledge analyses stratified by gist and verbatim knowledge. However, this analysis was rather exploratory than confirmatory in nature. We analysed informed choice – based on risk knowledge, attitude towards and the intention to participate – as a dichotomous dependent variable in binary logistic regression models. Another outcome variable was the 3-item graph rating scale. To control for type I and II errors, we chose a conventional alpha level of .05 and a beta-level of .20. All statistical computing and graphs were carried out with the language R version 3.5.3.

Results

We present the results of the three research strands separately. First, briefly the results of the pre-tests, then the online experiment, and afterwards the results of the qualitative, complementary evaluation. The quantitative and qualitative results of part two and three were mixed on the level of interpretation and conclusion afterwards.

Results of the Pre-Tests

In January 2017, we conducted a pilot test to ensure the feasibility at the HAW usability laboratory with one participant. The questionnaire was too exhausting in terms of scope, especially the graphic literacy scale. From an expert (Yasmina Okan) we received an unvalidated short version of the scale, which was validated later (Okan et al., 2019). We conducted the first pre-test with six participants pairwise. The content analysis revealed minor spelling and format issues. More severe problems were related to comprehension and excessive demand, mainly by the visualisation showing the results of screening. We simplified this graph and rephrased confusing sentences and questions. The second pre-tests took place at the participants' homes with seven participants, three in single and four pairwise sessions. Further minor issues occurred with unclear terms, phrases, and annotations. All major prior problems had been solved and the newly discovered minor problems were adjusted.

Table 1. Participation, Completion, and Dropout Overall and per Recruitment Method

	Experimental Groups								Full Sample	
	Text-Only		StaticVis		InfoVis		InfoVis+			
	n	%	*n*	%	*n*	%	*n*	%	*N*	%
Participation overall										
Started	151	100	147	100	149	100	152	100	599	100
Completed	96	63.6	85.0	57.8	75	57.1	76	50.0	342	57.1
Dropped-out	55	36.4	62.0	42.2	64	42.9	76	50.0	257	42.9
Participation panel sample										
Started	64	100	63.0	100	63	100	64	100	254	100
Completed	52	81.2	51.0	81.3	46	73.9	48	75.4	197	77.6
Dropped-out	12	18.8	12.0	18.7	17	26.1	16	24.6	57	22.4
Participation convenience sample										
Started	87	100	84.0	100	86	100	88	100	345	100
Completed	44	50.6	34.0	40.5	39	45.4	28	31.8	145	42.0
Dropped-out	43	49.4	50.0	59.5	47	54.6	60	68.2	200	58.0

Results of the Online Experiment

Between September 2017 and April 2018, 599 persons started the online experiment, 342 participants finished it, and 257 dropped out. Four more were excluded because they were younger than 30 years or older than 49 years. The overall dropout rate was 42.9% (Table 1). The final sample comprised of 338 participants. Participants were on average 39 years old. About a quarter achieved a higher education entrance certificate and more than a third a university degree (Table 2). Most used a computer (72.8%) to conduct the experiment, with others using tablets, mobiles, or other devices. About two thirds reported knowing someone with breast cancer, some had undergone a mammogram, or had searched for screening information (Table 2). The mean numeracy score was 2.3, the graph literacy score 2.1, and the mean duration 18 minutes. Participants with the additional tasks needed a bit longer (M = 19.8 minutes, Table 3). We tested all parameters for differences between the experimental groups and could not detect any statistically significantly differences.

H1 was that InfoVis, and additional reflective tasks would support risk knowledge and scores would be higher in these groups. The overall mean risk knowledge score was 7.6 (SD = 2, 95% CI = 7.4 to 7.8; single item results in the Appendix, Appendix A). In the one-way ANOVA risk knowledge was not statistically significantly different between all groups, $F(23, 334) = 1.92, p = .13, \eta^{2} = .13$. We conducted planned contrasts ANOVA to compare Text-only, StaticVis, InfoVis, and InfoVis+ groups described previously (see Statistical Methods), and ANCOVA to integrate moderator variables in a stepwise approach. In the final ANCOVA model, numeracy and graph literacy as well as their interaction term remained statistically significant predictors for risk knowledge, but no other variables (Table 4). The StaticVis group showed statistically significantly better risk knowledge in contrast to InfoVis and InfoVis+ groups combined with a small effect size. The post hoc Tukey test comparing all groups did not show any statistically significantly differences.

Table 2. Number, Education, Prior Information, and Applied Device of Participants

Participants	Experimental Groups								Full Sample	
	Text-Only		StaticVis		InfoVis		InfoVis+			
	n	%	*n*	%	*n*	%	*n*	%	*N*	%
	96	28.0	85	24.9	82	24.9	75	22.2	338	100
Highest educational level										
Secondary school certificate ("Hauptschule")	6	6.3	4	4.7	0	0.0	3	4.0	13	3.9
Secondary school certificate ("Realschule")	22	22.9	17	20.0	17	20.7	19	25.3	75	22.2
Polytechnic secondary school certificate	1	1.0	1	1.2	2	2.4	5	6.7	9	2.7
Advanced technical college certificate	5	5.2	8	9.4	9	11.0	4	5.3	26	7.7
Higher education entrance certificate ("Abitur")	23	24.0	18	21.2	22	26.8	21	28.0	84	24.9
University degree	37	38.5	35	41.2	30	36.6	21	28.0	123	36.4
Other	2	2.1	2	2.4	2	2.4	2	2.7	8	2.4
Prior information										
Prior mammography	24	25.0	16	18.8	20	24.4	13	17.3	73	21.6
Prior information-seeking	29	30.2	23	27.1	23	28.1	19	25.3	94	27.8
Knowing a person with breast cancer	65	67.7	61	71.8	57	69.5	48	64.0	231	68.3
Applied Device										
Computer	66	69.5	68	80.0	62	75.6	50	66.7	246	72.8
Tablet	11	12.0	4	4.7	7	8.5	11	14.7	33	9.8
Mobile	16	16.8	13	15.3	12	14.6	14	18.7	55	16.3
Other	3	3.1	0	0.0	1	1.2	0	0.0	4	1.1

According to H2, we assessed for differences in verbatim and gist knowledge, again without statistically significant differences (Table 5).

Regarding informed choice, most participants intended to do the screening, about two thirds had a positive attitude and more than half had sufficient knowledge (Table 6). An informed decision was made by 40.4%. In the simple logistic regression model with the Text-only group as reference, there was no statistically significant effect. To test H3, we ran multivariate logistic regression models including effects of the presumed moderating variables on informed choice. The final model showed no differences in informed choice between groups, with only numeracy as a significant predictor of informed choice (Table 7). Goodness-of-fit measures for

Table 3. Age, Numerical and Graphical Skills of Participants

	Experimental Groups								Full Sample	
	Text-Only		StaticVis		InfoVis		InfoVis+			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years, range: 30-49)	39.5	5.7	39.0	5.3	39.5	6.0	39.4	6.2	39.4	5.8
Duration (minutes)	17.7	7.1	17.5	7.3	17.2	6.9	19.8	7.3	18.0	7.2
Numeracy (score: 0-6)	2.3	1.7	2.1	1.6	2.2	1.6	2.6	1.8	2.3	1.7
Graph literacy (score: 0-4)	2.1	1.1	2.1	1.1	2.2	1.1	2.1	1.1	2.1	1.1

the model were poor (Hosmer and Lemeshow's R^2 = .03, Cox and Snells R^2 = .05, and Nagelkerke's R^2 = .06). Graph rating did not differ statistically significantly between the StaticVis, InfoVis, and InfoVis+ group. The mean sum score was 9.01. No effect modifying variable was associated with graph rating and thus none included in further analyses. Because of skewed distribution we conducted robust Kruskal-Wallis rank sum test, which did not reveal a significant effect between the groups (χ^2 (2) = 1.91, p = .385). Regarding participation and dropout rates, we observed a substantial difference between the experimental groups (Table 1). In the InfoVis+ group, 13.6% fewer participants finished the experiment than in the Text-only group (50% vs. 63.6%). Statistically, dropout rates were not significantly different between all groups (χ^2 (3) = 5.74, p = .120). Nevertheless, we assumed that recruitment methods might have biased participation and stratified accordingly: into participants recruited by the web-based panel provider and those via convenient methods. The difference in dropout rates between these sample were statistically significant (22.4% vs. 58%; χ^2 (3) = 73.9, p < .001). More participants of the convenience sample were well educated (χ^2 (6) = 41.8, p < .001), more knew people with breast cancer (χ^2 (1) = 13.31, p < .001), had higher numeracy (M = 2.6 vs. 2; t (283.12) = 3.26, p = .001) and graph literacy scores (M = 2.4 vs. 1.9; t(271.04) = 4.57, p = .001). Attitude towards mammography screening (M = 9.35 vs. 7.86, t(307.92) = - 4.27, p < .001) and intention to participate was higher (88.3% vs. 78%, χ^2 (1) = 5.76, p = .016) in the panel sample. However, differences in risk knowledge results were not statistically significant (M = 7.8 vs. 7.4, t(335.93) = 1.79, p = .073), neither verbatim nor gist knowledge results. To assess the impact of these differences, we repeated the statistical analyses stratified per sample. In the ANCOVA model with risk knowledge as outcome, the effect favouring the StaticVis to the two interactive groups was not significant in each sample, while the size of the interaction term (r = .17, p = .018). Additionally, in this sample the odds of an informed choice were statistically significantly lower in the InfoVis group compared to Text-only in the logistic

Table 4. Results of ANCOVA with Planned Contrasts for Risk Knowledge

Predictor Variables	*t* (6, 331)	*r*	*95% CI*		*p*
			LL	OL	
Numeracy	4.48	.24	0.14	0.33	< .001
Graph literacy	2.31	.13	0.02	0.23	.02
Interaction: Numeracy x Graph literacy	-2.37	-.13	-0.23	-0.02	.02
Planned contrasts					
Text-only vs. StaticVis/InfoVis/InfoVis+	-1.38	-.08	-0.18	0.03	.17
StaticVis vs. InfoVis/InfoVis+	-2.04	-.11	-0.21	0.00	.04
InfoVis vs. InfoVis+	0.05	.82	-0.06	0.15	.40

Table 5. Means, Standard Deviations, and One-Way Analyses of Variance for Risk Knowledge and Attitude

Continuous Outcome Variables	Experimental Groups									
	Text-Only		StaticVis		InfoVis		InfoVis+			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	$F(3, 334)$	η^2
Risk knowledge (score: 0-12)	7.8	2.0	7.8	2.0	7.1	2.2	7.6	2.0	1.92	.13
Verbatim knowledge (score: 0-5)	2.6	1.1	2.6	1.1	2.3	1.2	2.5	1.2	1.38	.11
Gist knowledge (score: 0-5)	3.7	1.0	3.7	1.0	3.5	1.6	3.8	1.1	0.84	.08
									$F(3, 328)$	η^2
Attitude (score: 0-16)	9.1	3.3	8.5	3.1	9.0	3.6	8.2	3.0	1.28	.11

interaction term ($r = .17$, $p = .018$). Additionally, in this sample the odds of an informed choice were statistically significantly lower in the InfoVis group compared to Text-only in the logistic regression model ($OR = 0.38$, $p = .030$). These effects were not evident in the convenience sample.

Results of the Qualitative Content and Relational Analysis

The qualitative data for the third research strand were derived from think-aloud protocols and focused interviews, recorded in seven individual sessions between January and April 2018. After an introduction the participants worked through the decision aid. The investigator motivated thinking aloud and was approachable but did not provide explanations or help. Mostly the InfoVis and tasks attracted attention and were further questioned. Only rarely the investigator applied a more directive approach based on prepared structured questions towards the ending of the interviews. Sessions lasted between 28 and 64 minutes.

The final set of the content analysis comprised nine themes: cognitive load, cognitive dissonance, feeling tested, attitude and experience, rating, decisional balancing, decision, interviewer effects, and miscellaneous issues. Three themes had no and the others up to seven subcategories. We coded all interviews accordingly. Appendix B presents all themes, subcategories, and quotes to illustrate the coding. We created a relational map to depict and interpret the themes in their relation to the experimental stimuli. Figure 2 shows how, and which themes and codes are linked to the InfoVis, reflective tasks, and the intention whether to participate in the screening. We neglected the first two InfoVis because no participant related them to decision-making; likewise the themes interviewer effect, miscellaneous issues and rating are not related to the stimuli and thus not integrated.

The InfoVis and tasks affected cognitive load varying on the complexity of the underlying statistical concepts and personal abilities. The application was not intuitive and fast at the beginning. Most participants showed that they needed some time to handle, understand, and interpret the InfoVis. Later, most participants utilised them appropriately. The InfoVis showing

Table 6. Absolute and Relative Results of Informed Choice Dimensions and Informed Choice

	Experimental Groups	Full Sample

	Text-Only		StaticVis		InfoVis		InfoVis+			
	n	%	*n*	%	*n*	%	*n*	%	*N*	%
Intention to participate (*yes*)	84	87.5	70	82.4	69	84.2	61	81.3	284	84
Sufficient knowledge (>= 6)	59	61.5	50	58.9	36	44.0	42	56.0	187	55.3
Positive attitude (>= 8)	63	67.0	55	65.5	53	66.3	47	63.5	218	65.7
Informed choice (*yes*)[a]	40	42.6	37	44.0	27	33.8	30	40.5	134	40.4

Note. [a] Informed choice is based on sufficient knowledge combined with either a positive attitude and intention to participate or a negative attitude and no intention.

the mortality benefit did not pose extra cognitive effort and was consistently interpreted correctly. In contrast, the overdiagnosis InfoVis was cumbersome. Participants had difficulties to transfer the previously given explanation to the interactive icon-arrays and the task. One participant processed the InfoVis and task correctly despite a misinterpretation of overdiagnosis. Another one realised that she previously misinterpreted overdiagnosis and came to a better understanding. However, most were confused and bothered.

Cognitive dissonance includes reactions such as surprise, uncertainty, or other negative emotions due to information that did not match the participants' expectations. Most were surprised by the seemingly small mortality benefit. Regarding overdiagnosis, the exact number was less important than the fact that can occur at all. Participants criticised that this disadvantage was overemphasised. The risk-benefit ratio of the screening was perceived as negative. The reflective tasks evoked the feeling of being tested in several participants with accompanying reactions of annoyance, stress, or embarrassment. Some noted that the tasks helped to keep the numbers in mind, and some ignored the tasks completely. The attitudes and experiences towards risk, breast cancer, mammography, and health in general shaped participants' decision-making. Disease histories of friends, family and personal experience with health providers were crucial. Two participants reported critical previous graph experiences. Others stated they were inexperienced with graphs and the handling was cumbersome. Some mentioned the graphs caught attention and were helpful to grasp statistical information. Comments and ratings concerning the InfoVis were mixed. Some features like colours or shape of icons were highlighted, while others were criticised, like too much information in pop-ups.

Table 7. Results of Multivariate Logistic Regression for Informed Choice (*yes*) [a]

Predictor Variables	*b*	*SE*	*OR*	95% CI		*p*
				LL	OL	
Numeracy	-0.25	0.07	1.29***	1.13	1.48	< .001
Experimental groups						
Text-only (reference)						
StaticVis	0.09	0.31	1.1	0.60	2.03	.75
InfoVis	-0.39	0.32	0.68	0.36	1.27	.22
InfoVis+	-0.19	0.32	0.82	0.44	1.56	.56

Note. [a] Informed choice is based on sufficient knowledge combined with either a positive attitude and intention to participate or a negative attitude and no intention. ****p* < .001.

Regarding decision-making, four participants reported they were determined before, and the information did not change their decision. The other three were weighing the pros and cons of the screening. Still, experiences, stories and attitudes played a major role in their decision-making.

Discussion

Our aim was to investigate whether and how InfoVis and reflective tasks can support risk communication and medical decision-making regarding mammography screening, based on a web-version of the current decision aid for the German mammography screening programme revised in 2017, which reached more than three million women in 2021 (Hand, 2023). We hypothesised that interactive processing would improve risk knowledge, which we operationalised as gist and verbatim knowledge to investigate if the fuzzy trace theory can

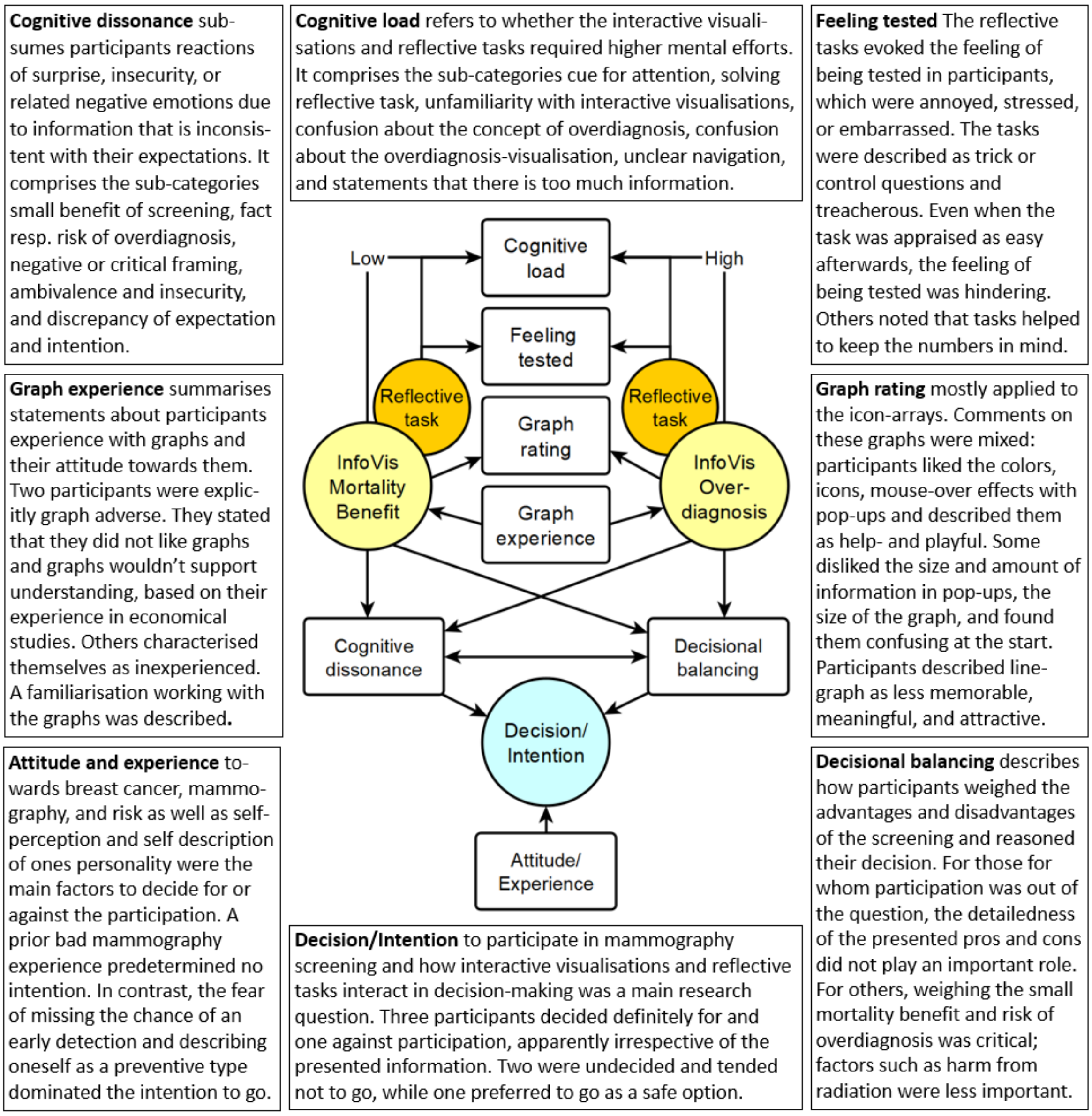

Figure 2. Relational Analysis with Descriptions of Themes

explain expected effects. Further, we investigated effects on informed choice and applied qualitative methods to explore the experience and processing of interactive presented risk information and reflective tasks.

Risk Knowledge and Informed Choice

The analyses of the online experiment did not show statistically significant differences regarding risk knowledge or informed choice. A subsequent analysis showed even significantly inferior risk knowledge (Table 4). However, all effect sizes were small. The qualitative analysis showed distinct processing of risk information and understanding comparing the two equally formatted InfoVis representing mortality benefit and overdiagnosis. The first was easy to grasp and the task easy to solve, while the latter posed problems for most participants due to difficulties understanding the concept of overdiagnosis.

In controlled experiments, Ancker et al. (2011) did not detect differences regarding risk estimates, risk feelings, or intention, and Zikmund-Fisher et al. (2011) reported mostly no differences albeit partly worse performance with interactive icon-arrays. Our results are in line with these studies, but do not support the results by Okan et al. (2015) that reflective tasks improve risk understanding. However, the latter study investigated undergraduate students, which probably differ from our sample in many aspects. Our participants were older and all females. We suggest that the stereotype-threat might have interfered the processing of information and performance, based on the qualitative results. The stereotype-threat describes the phenomenon that women underperform in maths tests because of the stereotype that women have weaker maths abilities and associated discomfort (Aronson et al., 1998; Jones, 2017; Spencer et al., 1999). We assume that our category "feeling tested" relates to this phenomenon and interferes with participants performance (Figure 2). In students, like in the Okan study, the gender stereotype threat appears diminished or absent (Flore et al., 2019; Okan, Garcia-Retamero, et al., 2015; Picho et al., 2013).

Compared to the evaluation of the original decision aid, the results of the three adapted knowledge items (s. Appendix; Risk Knowledge Scale, items 7, 11 and 12) were consistent in women younger than 50 regarding breast cancer risk and age (46% vs. 47.3%), screening results (85% vs. 83.1%), and the meaning of overdiagnosis (65% vs. 65.4%) (IQWIG, 2017). The German survey by Berens and colleagues (2015) reported less participants with sufficient knowledge (31.5% vs. 55.3%) and an informed choice (27.1% vs. 40.4%) than in our study. Three single, comparable knowledge items reflected this and showed fewer survey participants correctly assessed the likelihood of a true positive result (67.7% vs. 96.5%), the likelihood of a false negative result (52.1% vs. 89.3%), and the likelihood of diagnosis with or without screening (18.9% vs. 47.6%) (Berens et al., 2019). Several differences complicate this comparison: the survey participants were older and actual invitees, they received a mammography leaflet that was revised when our study was conducted, Turkish immigrant women were oversampled, and other knowledge items differed. Reder and Kolip (2017) conducted a randomised study to test a mammography decision aid with similar participants, measures, and material. Adequate knowledge and informed choice were higher compared to usual care (leaflet) at first post assessment (2 weeks; 66.8% vs. 31.4%; resp. 61.5% vs. 28.9%), and comparable to our results at the second assessment (3 month; 51.3% vs. 55.3%; resp. 39.8% vs. 40.4%). But the later assessment and unavailability of the tested decision aid complicate this comparison too, additional to the aforementioned differences.

Other informed choice domains—attitude and intention—varied between groups but did not differ statistically (Table 6). Both were higher in the panel than in the convenient sample, and some variation might be due to sampling strategies. The higher attrition mainly in the additional task group might hint to a negative effect of the tasks on attitude. However, without statistical significance random variation might be attributable too. In the qualitative interviews, some interviewees reported that the decision aid did not change their attitude or intention towards mammography screening, while others reported that the information was relevant for their decision. The motives for or against the screening were diverse and intricate.

InfoVis and reflective tasks may influence decision-making by affecting decisional-balancing and cognitive dissonance while increasing cognitive load might complicate information processing (Figure 2). Mason and colleagues observed that nearly all participants in a qualitative study referred to graphical risk information, but the processing with graphs varied widely between individuals, like in our study (Mason et al., 2014). Numerical and graphical skills seem relevant, but methodological issues like sampling and observation effects must be considered, when comparing quantitative and qualitative results.

Motives like security, attitude towards risk, experiences and stories about breast cancer had a major impact. Other authors highlighted that emotions play an important part in decision-making in screening, and cognitive comprehension has only little effect (Johansson & Brodersen, 2015). Qualitative research has shown that topics like reassurance, attitudes to risk, trust or scepticism, and others are important perspectives of women on breast cancer screening (Hersch et al., 2016). Further, the nature and extent of knowledge regarding screening decisions is difficult to determine, and potentially relevant information hard to define. It can range from all information available to omitting risk information. Criticism on the quantitative imperative emphasises that risk information is difficult to understand for most and may not always be adequate. It is also under debate, how to determine thresholds for adequate knowledge regarding informed choice (Ghanouni et al., 2016; Hawkes, 2012; P. H. Schwartz, 2011). The measures we applied from prior research allow an evaluation, focussing on the quantitative features we investigated, but the highlighted disadvantages must be considered.

Evaluation Theory

We assumed gist knowledge to be higher with InfoVis and verbatim knowledge lower compared to textual information. Overall, gist knowledge was statistically significantly higher than verbatim ($M = 3.7$ vs. 2.5, $t(668.07) = 13.56$, $p < .05$), but did not differ significantly between any groups in any analyses. There was no evidence that InfoVis specifically supported gist knowledge. The qualitative evaluation underpins this finding. Fuzzy, gist-based processing with InfoVis should have been intuitive, effortless, and fast, which was rarely the case. The InfoVis depicting mortality benefits seemed easy and intuitive for all, but especially the InfoVis about overdiagnosis was difficult and confusing for most. Processing depended on whether the user was familiar with InfoVis and whether the underlying concept of the data was simple and intuitive or complex. Unfamiliar formats, complex data concepts and reflective tasks addressed verbatim processing.

Our fuzzy trace approach seemed unsuitable because neither the InfoVis nor the tasks specifically targeted the gist system. Meanwhile, cognitive load was an important issue, and the cognitive load theory might be more appropriate to investigate InfoVis applications. Intrinsic cognitive load refers to the load that is inherent to the information itself, while extraneous load is the extra effort that is due to the design of information (Sweller, 2011). Like Mason and Ancker and colleagues we found evidence that InfoVis can draw attention to quantitative information (Ancker et al., 2011; Mason et al., 2014). With low intrinsic cognitive load InfoVis are easy to grasp, while a high intrinsic load poses problem. For women in our study, processing of mortality benefit information with InfoVis was suitable and with low cognitive load. Other risk information like overdiagnosis seem easier to grasp with text only and without interactivity to avoid high intrinsic load. Cognitive load measures might help to identify data concepts that are appropriate for InfoVis. Additional reflective tasks might evoke negative feelings and cause attrition in women but might work in younger and male persons. What combinations of textual, graphical, and interactive risk information formats work best for specific target groups could be addressed in further studies, aiming at low cognitive load.

Limitations

We did not test with actual mammography screening invitees, but with younger, possible future users. We calculated the sample size only for an overall effect on risk knowledge in the ANOVA, not for the contrasted analyses or informed choice. With more statistical power effects might have become significant. However, the effect sizes are small. The completion rates differed, most impaired by the reflective tasks in the convenience sample. Associated increased cognitive load and the feeling of being tested with sometimes strong negative emotional reactions complicated processing according to the qualitative results. Technical issues may have contributed in the online experiment. Significantly fewer participants in the convenience sample conducted the experiment on a computer screen. The interactive web-based formats were not pre-tested on mobile devices and usability may have been impaired. Further, the motivation to participate might have differed. Panel members are familiar with web-based studies and compensation probably strengthened motivation to complete the experiment. Convenience sample members attended anonymously without compensation and could quit without consequences. Hence differential motivation depending on recruitment, the stereotype threat and usability issues regarding applied devices appeared to affect completion and impaired internal validity.

Conclusion

We aimed to clarify whether InfoVis and reflective tasks help lay users to better understand risks about mammography screening, support informed decision-making, how risk information is processed, and how users experience InfoVis and tasks. In an online experiment we did not find evidence that interactive icon-arrays can convey risks better than text or static graphs or affect informed choice. We operationalised the fuzzy-trace theory but could not detect different processing based on gist and verbatim knowledge. Qualitative results showed that visualisations can distort processing and address verbatim processing when the underlying risk concept is complex and elusive like overdiagnosis. However, when InfoVis present easier concepts like a mortality benefit, they are comprehensible and suitable to emphasise this information. However, usage should be limited and not substantially alter cognitive efforts. We suggest the cognitive load theory as promising to identify appropriate risk information for InfoVis and their evaluation. Pre-testing, especially on mobile devices, should be mandatory to avoid usability issues and investigate cognitive load. Reflective tasks do not seem advisable for users that may be prone to the stereotype threat, like the women in this study. However, our findings should be replicated with actual mammography invitees, and future research should investigate, for which other situations and target groups InfoVis and reflective tasks might be suitable to support textual and graphical health risk information.

Notes

1. https://infogram.com/.

Ethical Approval

The research project was approved by the ethics committees of the University of the West of Scotland (UWS) and the Hamburg University of Applied Sciences (HAW, Germany).

Funding

The research was supported by funding from the Hamburg University of Applied Sciences to finalise the online experiment by an online panel. Overall, the project was financed by the main author as required for his doctoral thesis.

Conflict of Interest

There are no conflicts of interest.

References

Alexander, L. K., Lopes, B., Richetti-Masterson, K., & Yeatts, K. B. (2015). *ERIC notebook* (2nd ed.). Epidemiologic Research and Information Center (ERIC). https://sph.unc.edu/files/2015/07/nciph_ERIC10.pdf

Alroobaea, R., & Mayhew, P. J. (2014). How many participants are really enough for usability studies? In *2014 Science and Information Conference* (pp. 48–56). https://doi.org/10.1109/SAI.2014.6918171

Ames, A. G., Metcalfe, S. A., Archibald, A. D., Duncan, R. E., & Emery, J. (2015). Measuring informed choice in population-based reproductive genetic screening: A systematic review. *European Journal of Human Genetics*, *23*(1), 8–21. https://doi.org/10.1038/ejhg.2014.89

Amidon, T. R., Nielsen, A. C., Pflugfelder, E. H., Richards, D. P., & Stephens, S. H. (2021). Visual risk literacy in "flatten the curve" COVID-19 visualizations. *Journal of Business and Technical Communication*, *35*(1), 101–109. https://doi.org/10.1177/1050651920963439

Ancker, J. S. (2020). The COVID-19 pandemic and the power of numbers. *Numeracy*, *13*(2), Article 2. https://doi.org/10.5038/1936-4660.13.2.1358

Ancker, J. S., Weber, E. U., & Kukafka, R. (2011). Effects of game-like interactive graphics on risk perceptions and decisions. *Medical Decision Making*, *31*(1), 130–142. https://doi.org/10.1177/0272989X10364847

Anthony, A. A., & Joseph, M. T. (2014). *SMEs in Indian textiles: The impact of globalization in a developing market*. Palgrave Macmillan. https://doi.org/10.1057/9781137444578_14

Aro, A. R., de Koning, H. J., Absetz, P., & Schreck, M. (1999). Psychosocial predictors of first attendance for organised mammography screening. *Journal of Medical Screening*, *6*(2), 82–88. https://doi.org/10.1136/jms.6.2.82

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice* (pp. 83–103). Elsevier. https://doi.org/10.1016/B978-012679130-3/50039-9

Berens, E.-M., Kaucher, S., Van Eckert, S., Reder, M., Kolip, P., & Spallek, J. (2019). Knowledge about mammography screening in Germany by education and migrant status – results of a cross-sectional study (InEMa). *Applied Cancer Research*, *39*(1), 6. https://doi.org/10.1186/s41241-019-0076-1

Berens, E.-M., Reder, M., Kolip, P., & Spallek, J. (2014). A cross-sectional study on informed choice in the mammography screening programme in Germany (InEMa): A study protocol. *BMJ Open*, *4*(9), Article e006145. https://doi.org/10.1136/bmjopen-2014-006145

Berens, E.-M., Reder, M., Razum, O., Kolip, P., & Spallek, J. (2015). Informed choice in the German mammography screening program by education and migrant status: Survey among first-time invitees. *PLOS ONE*, *10*(11), Article e0142316. https://doi.org/10.1371/journal.pone.0142316

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making*, *35*(4), 539–557. https://doi.org/10.1177/0272989X14547740

Bonner, C., Trevena, L. J., Gaissmaier, W., Han, P. K. J., Okan, Y., Ozanne, E., Peters, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current best practice for presenting probabilities in patient decision aids: Fundamental principles. *Medical Decision Making*, *41*(7), 821-833. https://doi.org/10.1177/0272989X21996328

Busch, C., De Maret, P. S., Flynn, T., Kellum, R., Le, S., Meyers, B., Saunders, M., White, R., & Palmquist, M. (2005). *Content Analysis*. Colorado State University. https://writing.colostate.edu/guides/guide.cfm?guideid=61

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann Publishers.

Carpendale, S. (2008). *Information visualization: Human-centered issues and perspectives*. Springer.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. SAGE Publications.

Das Mammographie Screening-Programm. (n.d.). *Informationsmaterial English*. Kooperationsgemeinschaft Mammographie. Retrieved from https://www.mammo-programm.de/de/informationsmaterial-english

Elmore, J. G. (2016). Breast cancer screening: Balancing evidence with culture, politics, money, and media. In N. Houssami & D. Miglioretti (Eds.), *Breast cancer screening* (pp. 1–27). Elsevier. https://doi.org/10.1016/B978-0-12-802209-2.00001-2

Etnel, J. R. G., de Groot, J. M., El Jabri, M., Mesch, A., Nobel, N. A., Bogers, A. J. J. C., & Takkenberg, J. J. M. (2020). Do risk visualizations improve the understanding of numerical risks? A randomized, investigator-blinded general population survey. *International Journal of Medical Informatics*, *135*, Article 104005. https://doi.org/10.1016/j.ijmedinf.2019.104005

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: Ten steps to better risk communication. *JNCI Journal of the National Cancer Institute*, *103*(19), 1436–1443. https://doi.org/10.1093/jnci/djr318

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Flore, P. C., Mulder, J., & Wicherts, J. M. (2019). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology*, 1–35. https://doi.org/10.1080/23743603.2018.1559647

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 18–30). Springer. http://link.springer.com/10.1007/3-540-46037-3_3

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A.-S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, *31*(3), 286–296. https://doi.org/10.1037/a0024850

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989X10373805

Galesic, M., & Garcia-Retamero, R. (2013). Graph literacy for health. In R. Garcia-Retamero & M. Galesic (Eds.), *Transparent communication of health risks* (pp. 53–65). Springer. http://link.springer.com/10.1007/978-1-4614-4358-2_4

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors*, *59*(4), 582–627. https://doi.org/10.1177/0018720817690634

Garcia-Retamero, R., Cokely, E. T., Ghazal, S., & Joeris, A. (2016). Measuring graph literacy without a test: A brief subjective assessment. *Medical Decision Making*, *36*(7), 854–867. https://doi.org/10.1177/0272989X16655334

Garcia-Retamero, R., & Galesic, M. (2010). Who proficts from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine*, *70*(7), 1019–1025. https://doi.org/10.1016/j.socscimed.2009.11.031

Garcia-Retamero, R., Okan, Y., & Cokely, E. T. (2012). Using visual aids to improve communication of risks about health: A review. *The Scientific World Journal*, *2012*(1), Article 562637. https://doi.org/10.1100/2012/562637

Gemeinsamer Bundesausschuss. (2014). *Konkretisierung des Auftrags des Gemeinsamen Bundesausschusses an das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen: Erstellung eines Einladungsschreibens und einer Versicherteninformation zum Mammographie-Screening*. https://www.g-ba.de/downloads/40-268-2923/2014-08-21_IQWiG-Beauftragung_Versicherteninformation_Mammographie-Scr_Konkretisierung.pdf

Gemeinsamer Bundesausschuss. (2017). *Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Krebsfrüherkennungs-Richtlinie (Mammographie-Screening): Einladungsschreiben und Entscheidungshilfe sowie weitere Änderungen*. https://www.g-ba.de/downloads/40-268-4504/2017-07-20_KFE-RL_Mammographie-Screening-Einladungsschreiben-Entscheidungshilfe_TrG.pdf

Ghanouni, A., Renzi, C., Meisel, S. F., & Waller, J. (2016). Common methods of measuring 'informed choice' in screening participation: Challenges and future directions. *Preventive Medicine Reports*, *4*, 601–607. https://doi.org/10.1016/j.pmedr.2016.10.017

Gigerenzer, G. (2014). Breast cancer screening pamphlets mislead women. *BMJ*, *348*, Article g2636. https://doi.org/10.1136/bmj.g2636

Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *JNCI Journal of the National Cancer Institute*, *101*(17), 1216–1220. https://doi.org/10.1093/jnci/djp237

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hawkes, N. (2012). „Citizens' jury" disagrees over whether screening leaflet should put reassurance before accuracy. *BMJ*, *345*, Article e8047. https://doi.org/10.1136/bmj.e8047

Hersch, J., Jansen, J., Irwig, L., Barratt, A., Thornton, H., Howard, K., & McCaffery, K. (2011). How do we achieve informed choice for women considering breast screening? *Preventive Medicine*, *53*(3), 144–146. https://doi.org/10.1016/j.ypmed.2011.06.013

Hersch, J., Jansen, J., & McCaffery, K. (2016). Informed and shared decision making in breast screening. In N. Houssami & D. Miglioretti (Eds.), *Breast cancer screening* (pp. 403–420). Elsevier. https://doi.org/10.1016/B978-0-12-802209-2.00016-4

IQWIG. (2017). *Einladungsschreiben und Entscheidungshilfe zum Mammographie-Screening*. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWIG).

Isenberg, P., Zuk, T., Collins, C., & Carpendale, S. (2008). Grounded evaluation of information visualizations. In E. Bertini, A. Perer, C. Plaisant, & G. Santucci (Eds.), *Proceedings of the 2008 workshop on beyond time and errors: novel evaluation methods for information visualization*, Article 6. Association for Computing Machinery. https://doi.org/10.1145/1377966.1377974

Jenny, M. A., Keller, N., & Gigerenzer, G. (2018). Assessing minimal medical statistical literacy using the quick risk test: A prospective observational study in Germany. *BMJ Open*, *8*(8), Article e020847. https://doi.org/10.1136/bmjopen-2017-020847

Johansson, M., & Brodersen, J. (2015). Informed choice in screening needs more than information. *The Lancet*, *385*(9978), 1597–1599. https://doi.org/10.1016/S0140-6736(15)60258-6

Jones, P. R. (2017). Reducing the impact of stereotype threat on women's math performance: Are two strategies better than one? *Electronic Journal of Research in Education Psychology*, *9*(24), 587–616. https://doi.org/10.25115/ejrep.v9i24.1458

Kaltsa, A., Holloway, A., & Cox, K. (2013). Factors that influence mammography screening behaviour: A qualitative study of Greek women's experiences. *European Journal of Oncology Nursing*, *17*(3), 292–301. https://doi.org/10.1016/j.ejon.2012.08.001

Keller, C., & Siegrist, M. (2009). Effect of risk communication formats on risk perception depending on numeracy. *Medical Decision Making*, *29*(4), 483–490. https://doi.org/10.1177/0272989x09333122

Koo, K., Brackett, C. D., Eisenberg, E. H., Kieffer, K. A., & Hyams, E. S. (2017). Impact of numeracy on understanding of prostate cancer risk reduction in PSA screening. *PLOS ONE*, *12*(12), Article e0190357. https://doi.org/10.1371/journal.pone.0190357

Hand, E. (2023). *Jahresbericht Evaluation 2021. Deutsches Mammographie-Screening-Programm.* Kooperationsgemeinschaft Mammographie. *https://fachservice.mammo-programm.de/download/evaluationsberichte/Eval-2021-Webversion.pdf*

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Beltz Juventa.

Lagerlund, M., Hedin, A., Sparén, P., Thurfjell, E., & Lambe, M. (2000). Attitudes, beliefs, and knowledge as predictors of nonattendance in a Swedish population-based mammography screening program. *Preventive Medicine*, *31*(4), 417–428. https://doi.org/10.1006/pmed.2000.0723

Lagerlund, M., Sparén, P., Thurfjell, E., Ekbom, A., & Lambe, M. (2000). Predictors of non-attendance in a population-based mammography screening programme; socio-demographic factors and aspects of health behaviour. *European Journal of Cancer Prevention*, *9*(1), 25–33. https://doi.org/10.1097/00008469-200002000-00004

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*(5), 696–713. https://doi.org/10.1177/0272989X07307271

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs*, *25*, 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Malek, D., & Kääb-Sanyal, V. (2016). Implementation of the German mammography screening program (German MSP) and first results for initial examinations, 2005-2009. *Breast Care*, *11*(3), 183–187. https://doi.org/10.1159/000446359

Malloy-Weir, L. J., Schwartz, L., Yost, J., & McKibbon, K. A. (2016). Empirical relationships between numeracy and treatment decision making: A scoping review of the literature. *Patient Education and Counseling*, *99*(3), 310–325. https://doi.org/10.1016/j.pec.2015.10.002

Marteau, T. M., Dormandy, E., & Michie, S. (2001). A measure of informed choice. *Health Expectations*, *4*(2), 99–108. https://doi.org/10.1046/j.1369-6513.2001.00140.x

Mason, D., Boase, S., Marteau, T., Kinmonth, A. L., Dahm, T., Minorikawa, N., & Sutton, S. (2014). One-week recall of health risk information and individual differences in attention to bar charts. *Health, Risk & Society*, *16*(2), 136–153. https://doi.org/10.1080/13698575.2014.884544

Mathieu, E., Barratt, A. L., McGeechan, K., Davey, H. M., Howard, K., & Houssami, N. (2010). Helping women make choices about mammography screening: An online randomized trial of a decision aid for 40-year-old women. *Patient Education and Counseling*, *81*(1), 63–72. https://doi.org/10.1016/j.pec.2010.01.001

MAXQDA [Computer software]. (n.d.). *All-in-one qualitative & mixed methods data analysis tool.* MAXQDA. Retrieved from https://www.maxqda.com/

Merton, R. K., Lowenthal, M. F., & Kendall, P. L. (1990). *The focused interview: A manual of problems and procedures* (2nd ed). Collier Macmillan.

Michie, S., Dormandy, E., & Marteau, T. M. (2002). The multi-dimensional measure of informed choice: A validation study. *Patient Education and Counseling*, *48*(1), 87–91. https://doi.org/10.1016/S0738-3991(02)00089-7

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, *5*(2), 147–149. https://doi.org/10.1177/104973239500500201

Multmeier, J., Gaissmaier, W., & Wegwarth, O. (2014). Collective statistical illiteracy in health. In B. L. Anderson & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 39–58). Cambridge University Press.

Natter, H. M., & Berry, D. C. (2005). Effects of active information processing on the understanding of risk information. *Applied Cognitive Psychology*, *19*(1), 123–135. https://doi.org/10.1002/acp.1068

Nayak, J. G., Hartzler, A. L., Macleod, L. C., Izard, J. P., Dalkin, B. M., & Gore, J. L. (2016). Relevance of graph literacy in the development of patient-centered communication tools. *Patient Education and Counseling*, *99*(3), 448–454. https://doi.org/10.1016/j.pec.2015.09.009

Nelson, W., Reyna, V. F., Fagerlin, A., Lipkus, I., & Peters, E. (2008). Clinical implications of numeracy: Theory and practice. *Annals of Behavioral Medicine*, *35*(3), 261–274. https://doi.org/10.1007/s12160-008-9037-8

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In B. Arnold, G. van der Veer, & T. White (Eds.), *Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems* (pp. 206–213). Association for Computing Machinery. https://doi.org/10.1145/169059.169166

Okan, Y., Galesic, M., & Garcia-Retamero, R. (2015). How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making*, *29*(2-3), 271-294. https://doi.org/10.1002/bdm.1891

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: overcoming denominator neglect in risk comprehension: Individual differences in graph literacy. *Journal of Behavioral Decision Making*, *25*(4), 390–401. https://doi.org/10.1002/bdm.751

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied*, *21*(2), 178–194. https://doi.org/10.1037/xap0000045

Okan, Y., Janssen, E., Galesic, M., & Waters, E. A. (2019). Using the short graph literacy scale to predict precursors of health behavior change. *Medical Decision Making*, *39*(3), 183–195. https://doi.org/10.1177/0272989X19829728

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1). https://doi.org/10.1186/s41235-018-0120-9

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics*, *35*(1), 52–75. https://doi.org/10.1016/S1532-0464(02)00009-6

Payne, D. G., & Westerman, D. L. (2003). Research methods in cognition. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 365–388). Blackwell Publishing. http://doi.wiley.com/10.1002/9780470756973.ch17

Pérez-Montoro, Mario (2022). Visual communication in the management of a global health emergency. In P. Freixa, L. Codina M. Pérez-Montoro, & J. Guallar (Eds.), *Visualisations and narratives in digital media: Methods and current trends* (pp. 128-147). DigiDoc-EPI. https://doi.org/10.3145/indocs.2022.9

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, *153*(3), 299–333. https://doi.org/10.1080/00224545.2012.737380

Plaisant, C. (2004). *The challenge of information visualization evaluation*. In M. F. Costabile (Ed.), Proceedings of the working conference on Advanced visual interfaces (pp. 109-116). Association for Computing Machinery. https://doi.org/10.1145/989863.989880

Reder, M., Berens, E.-M., Spallek, J., & Kolip, P. (2019). Development of the informed choice in mammography screening questionnaire (IMQ): Factor structure, reliability, and validity. *BMC Psychology*, *7*(1), Article 17. https://doi.org/10.1186/s40359-019-0291-2

Reder, M., & Kolip, P. (2015). Does a decision aid improve informed choice in mammography screening? Study protocol for a randomized controlled trial. *BMC Women's Health*, *15*, Article 53. https://doi.org/10.1186/s12905-015-0210-5

Reder, M., & Kolip, P. (2017). Does a decision aid improve informed choice in mammography screening? Results from a randomised controlled trial. *PLOS ONE*, *12*(12), Article e0189148. https://doi.org/10.1371/journal.pone.0189148

Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, *28*(6), 850–865. https://doi.org/10.1177/0272989X08327066

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. *Judgment and Decision Making*, *7*(3), 332–359. https://doi.org/10.1017/S1930297500002291

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*, *31*(2–3), 180–206. https://doi.org/10.1016/j.dr.2011.07.004

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327

Riggin, L. J. C. (1997). Advances in mixed-method evaluation: A synthesis and comment. *New Directions for Evaluation*, *1997*(74), 87–94. https://doi.org/10.1002/ev.1074

Robinson, O. C. (2011). Relational analysis: An add-on technique for aiding data integration in qualitative research. *Qualitative Research in Psychology*, *8*(2), 197–209. https://doi.org/10.1080/14780887.2011.572745

Schaeffer, D., Haarmann, A., & Griese, L. (2023). *Professionelle Gesundheitskompetenz ausgewählter Gesundheitsprofessionen in Deutschland. Ergebnisse des HLS-PROF-GER*. Stiftung Gesundheitswissen. https://www.stiftung-gesundheitswissen.de/sites/default/files/brochure/pdf/2023_06_20_Ergebnisbericht_Studie_professionelle_Gesundheitskompetenz.pdf

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 69*(Suppl 2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966. https://doi.org/10.7326/0003-4819-127-11-199712010-00003

Schwartz, P. H. (2011). Questioning the quantitative imperative: Decision aids, prevention, and the ethics of disclosure. *Hastings Center Report*, *41*(2), 30–39. https://doi.org/10.1353/hcr.2011.0029

Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100. https://doi.org/10.1037/1076-898X.4.2.75

Shah, P. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 426–476). Cambridge University Press.

Smith, S. K., Trevena, L., Simpson, J. m., Barratt, A., Nutbeam, D., & McCaffery, K. J. (2010). A decision aid to support informed choices about bowel cancer screening among adults with low education: Randomised controlled trial. *BMJ: British Medical Journal*, *341*(7780), 1–13. psyh.

Someren, M. W. van, Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Spool, J., & Schroeder, W. (2001). Testing web sites: Five users is nowhere near enough. In M. (Mantei) Tremaine (Ed.), *Extended abstracts on human factors in computing systems* (pp. 285-286). Association for Computing Machinery. https://doi.org/10.1145/634067.634236

Stellamanns, J., Ruetters, D., Dahal, K., Schillmoeller, Z., & Huebner, J. (2017). Visualizing risks in cancer communication: A systematic review of computer-supported visual aids. *Patient Education and Counseling*, *100*(8), 1421–1431. https://doi.org/10.1016/j.pec.2017.02.003

Sterlingova, T., & Lundén, M. (2018). Why do women refrain from mammography screening? *Radiography*, *24*(1), e19–e24. https://doi.org/10.1016/j.radi.2017.07.006

Strech, D. (2014). Participation rate or informed choice? Rethinking the European key performance indicators for mammography screening. *Health Policy*, *115*(1), 100–103. https://doi.org/10.1016/j.healthpol.2013.11.012

Sweller, J. (2011). Cognitive load theory and e-learning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education* (pp. 5-6). Springer. https://doi.org/10.1007/978-3-642-21869-9_3

Trevena, L. J., Bonner, C., Okan, Y., Peters, E., Gaissmaier, W., Han, P. K. J., Ozanne, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current challenges when using numbers in patient decision aids: Advanced concepts. *Medical Decision Making*, 0272989X2199634. https://doi.org/10.1177/0272989X21996342

Trevena, L. J., Zikmund-Fisher, B. J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P. K., King, J., Lawson, M. L., Linder, S. K., Lipkus, I., Ozanne, E., Peters, E., Timmermans, D., & Woloshin, S. (2013). Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. *BMC Medical Informatics and Decision Making*, *13*(Suppl 2), Article S7. https://doi.org/10.1186/1472-6947-13-S2-S7

Van Weert, J. C. M., Alblas, M. C., van Dijk, L., & Jansen, J. (2021). Preference for and understanding of graphs presenting health risk information. The role of age, health literacy, numeracy and graph literacy. *Patient Education and Counseling*, *104*(1), 109–117. https://doi.org/10.1016/j.pec.2020.06.031

Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association for Information Systems*, *17*(7), 435–494. https://doi.org/10.17705/1jais.00433

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, *34*(4), 457–468. https://doi.org/10.1177/001872089203400407

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2011). Deceiving numbers: Survival rates and their impact on doctors' risk communication. *Medical Decision Making*, *31*(3), 386–394. https://doi.org/10.1177/0272989x10391469

Williamson, K. (2018). Populations and samples. In K. Williamson & G. Johanson (Eds.), *Research methods* (pp. 359–377). Elsevier. https://doi.org/10.1016/B978-0-08-102220-7.00015-7

Zikmund-Fisher, B. J., Dickson, M., & Witteman, H. O. (2011). Cool but counterproductive: Interactive, web-based risk communications can backfire. *Journal of Medical Internet Research*, *13*(3), Article e60. https://doi.org/10.2196/jmir.1665

Zikmund-Fisher, B. J., Mayman, G., & Fagerlin, A. (2014). Patient numeracy: What do patients need to recognize, think, or do with health numbers? In B. L. Anderson & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 80–104). Cambridge University Press. https://doi.org/10.1017/CBO9781139644358.005

Zikmund-Fisher, B. J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Kahn, V. C., Exe, N. L., Valerio, M., Holtzman, L. G., Scherer, L. D., & Fagerlin, A. (2014). Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, *34*(4), 443–453. https://doi.org/10.1177/0272989X13511706

Author Contributions

Conceptualisation (main idea, theory): Jan Stellamanns
Funding acquisition: Zita Schillmoeller (web panel)
Project administration: Jan Stellamanns
Methodology (design, operationalisation): Jan Stellamanns
Data collection: Jan Stellamanns
Data analysis: Jan Stellamanns
Writing – original draft: Jan Stellamanns
Writing – review & editing: Zita Schillmoeller & Keshav Dahal

Author Biographies

Jan Stellamanns is a PhD and scientific staff at the German Cancer Society. His area of expertise is health and risk communication.

Keshav Dahal is a professor of intelligent systems and the director of the Artificial Intelligence, Visual Communications and Network (AVCN) Research Institute at the University of the West of Scotland, UK. His research interests lie in the areas of applied AI, trust and security modelling in distributed systems, Blockchain technology and scheduling/optimisation problems.

Zita Schillmoeller is a professor of health science and statistical methods at the University of Applied Sciences in Hamburg (HAW). Her research expertise includes applied statistics and population-based evaluation, research methods, health education, children's health, and Public Health and violence.

Appendix

Appendix A. Single Risk Knowledge Items by Subscales Gist, Verbatim and Other Knowledge (Correct Answers)

Risk Knowledge	Experimental Groups								Full Sample	
	Text-Only		StaticVis		InfoVis		InfoVis+			
	n	%	*n*	%	*n*	%	*n*	%	*N*	%
Gist knowledge items										
Age and breast cancer incidence	75	78.1	62	72.9	62	75.6	57	76.0	256	75.7
Likelihood of a true positive screening result	92	95.8	84	98.8	77	93.9	73	97.3	326	96.5
Likelihood of false negative result or interval cancer	88	91.7	77	90.6	68	82.9	69	92.0	302	89.3
Likelihood of diagnosis with/without screening	42	43.8	38	44.7	39	47.6	42	56.0	161	47.6
Breast cancer mortality with/without screening	56	58.3	55	64.7	42	51.2	41	54.7	194	57.4
Verbatim knowledge items										
Age and breast cancer incidence	52	54.2	35	41.2	35	42.7	38	50.7	169	47.3
Likelihood of a true positive screening result	81	84.4	75	88.2	66	80.5	59	78.7	281	83.1
Likelihood of false negative result or interval cancer	65	67.7	65	76.5	50	61.0	42	56.0	222	65.7
Likelihood of diagnosis with/without screening	22	22.9	14	16.5	20	24.3	27	36.0	83	24.6
Breast cancer mortality with/without screening	29	30.2	32	37.7	16	19.5	22	29.3	99	29.3
Other knowledge items										
Frequency of positive screening results	56	83.3	55	76.5	42	72.0	41	69.3	256	75.7
Meaning of overdiagnosis	66	68.8	59	69.4	51	62.2	45	60.0	221	65.4

Appendix B. Results of Qualitative Interviews with Themes, Sub-Categories, and Example Quotes

Themes and Sub-Categories	Example Quotes
Cognitive load	
Cue for attention/ concentration	"Yes, to ask these questions then, because that always means you have to deal with the text again, but it's good that you can go back"; "... you tend to get stuck and maybe look again, then you go back and forth again to look"; "... I needed so much concentration for the graphics"
Solve reflective task	"Oh, what did we have, 15, 16, 17, 18, 19 ... I'll say 4"; "Oh no, 10 more. Look at that. I would have thought that, no, sure, that's 5 steps. 10 more."; "Oh, how many were there? I was just looking relatively. 10?"
Unfamiliarity/ Familiarisation with interactive InfoVis	"This one, right, that was the first one, you have to get into it first."; "… then you it takes a while to get into it. But I think you get it then."; "... not so difficult to understand, but not at first sight"
Confusion about the concept of overdiagnosis	"I don't understand that, for example, I don't understand that. So ... overdiagnoses …"; "Funny, how did I put that? These are ... I thought misdiagnoses ... that it's something that's not cancer."; "I didn't understand that at the beginning, but afterwards I understood that there are ... diagnoses that are meaningful and good, and some that ... are not meaningful."
Confusion about the interactive InfoVis about overdiagnosis	"Yeah, so this graph is kind of hard to understand, I think."; "With screening, without screening, 49 with, 59 without. So of the two tables, I wouldn't know which number you would like now."; "But I don't understand this. Why does this graph show the overdiagnosis? ... 10 less than those who took part in the screening? So, that's the ... then in case of doubt that's 10 wrong diagnoses, or what?"
Unclear navigation	"I didn't know where to go next ... oh, there's next."; "OK. Oh, now I must go back again, how can I go back?"; "What does that mean, press it? I don't understand that right now."
Too much information or text	"It's too detailed in some parts, I have the feeling."; "I think it's really a lot to read in any case, …"; "it was quite a lot, so I would have thought it was a bit less …"
Cognitive dissonance	
Small mortality benefit of screening	"So there's not that much difference now, is there?"; "And that the difference between those who go regularly and those who never go is … only 4. I thought that was quite crass."; "You get the feeling that it's not worth it to have a mammogram. So why should I go through the stress ... when there is only a difference of 4."
Fact/Risk of overdiagnosis	"And in retrospect, of course, there are many overdiagnoses, ne ... I'm rather afraid of these overdiagnoses."; "With the overdiagnosis, I find that super astounding, I wouldn't have expected that at all."
Unclarity/Discrepancy expectation – intention	"... the goal is to get more women to have mammograms, actually, or have I misunderstood?"; " ... because I think that if you invite the people, well, I assume that they want to invite the people to participate, ... so the disadvantages are also quite big."; "I asked myself all the time, what do you want ... what is the intention?"

Appendix B. Results of Qualitative Interviews with Themes, Sub-Categories, and Example Quotes (continued)

Themes and Sub-Categories	Example Quotes
Ambivalence and insecurity	"Maybe that's what irritates me, because somehow, I don't know what to do with it, how to categorise it. Should I go or should I not go?"; "... I just have the feeling that I'm drawing the wrong conclusions"; "... and you're a bit confused as to whether you should do it or not."
Negative/Critical framing of information	"Well, that's the only one where the text is illustrated by a graphic, but only in the negative example"; "I would emphasise the early detection, the positive aspects of it, and not the negative"; "And it has to be presented in a more positive way."
Adaptation/Reduce dissonance	"Absolutely. When I say that I'm surprised, it doesn't mean that I think it's stupid, mostly the opposite. Actually, I always find things that illuminate several sides good."
Feeling tested	"I really started sweating a bit. I thought, am I being tested or what? "; "At that moment, you just think it's more complicated than it really is, you know? The question was actually simple. But ... at the first moment you think, uh, test, what? I must think!"; "Oh God, now you should have remembered that?"; "I can make a fool of myself."; "Now surely comes another one of those control questions."
Attitude and experience	
Experience with and attitude towards breast cancer and mammography	"... this is the wife of a buddy. They've taken off both breasts... Imagine that! That's so bad."; "I'm only 42, but I've known people who've had something (breast cancer) at 35."; "... because I had a mammogram once and had informed myself a little bit"; "So I also know women who have done that (mammography) even in their late 30s."; "My mother has just told that once, so I know that a little bit."
Experience with and attitude towards graphs	"Numbers and charts, and I thought it was awful. I am not a numbers person."; "I'm not good at graphs, though, I have to say. I've always found graphs stupid."; "... and such a graph, because I'm probably not used to looking at graphs either"; "Otherwise, I think it's always good when it's presented in this way, i.e. in colour, simply because it's much easier to grasp than if the numbers were somehow written there."
Attitude towards risk	"If I were one of the four now, it would certainly have been worth it,"; "... to rule it out as far as possible. Or to also have the feeling, I have now done everything to ensure that it is covered"; “... then I would rather say, OK, let‘s take the chance“

Appendix B. Results of Qualitative Interviews with Themes, Sub-Categories, and Example Quotes (continued)

Themes and Sub-Categories	Example Quotes
Self-perception and personality	"My brain is like a sieve."; "I am also simply curious."; "But I'm also a preventative type."; "Unless you're at that age and you're worried"; "...that you can't just swallow and accept everything"; "... it' s a little bit my nature, that I always think, OK, what did you actually want?"
Decisional balancing	"But if you look at the statistics, you will rather say that it's not ... that if I don't participate, and if it's then diagnosed, if there are symptoms, that it's no longer curable. So the risk is lower, that's true, but it's not impossible."; "But if they actually have nothing in the past, then I would think that, oh, I always do my nodule tests at the gynaecologist, that's enough. I'm rather afraid of these overdiagnoses."
Decision	"Nope, it's a sure thing for me anyway."; "So I would definitely do that."; "But I would still do it."; "I think I would actually decide not to do it."; "I already decided against it 10 years ago."
Rating	
Rating the interactive icon-array InfoVis	"I find colours totally important … to just see ... how much it differs."; "That is actually quite cute made here with the girls (female icons)"; "... now this one, I find that very illustrative. The only thing is that it's too big. It should be more compact somehow ..."; "... it is playful, so you're more likely to get stuck and maybe look again ..."; "I think that's OK. So I don't think wow, great."; "It is somehow messy."
Rating the interactive line graph InfoVis	"For example, this graphic … with the two lines, I had somehow completely forgotten."; "I can't understand it by reading the above and just looking at it. Well, it's good that it's clickable again."; "I find it a bit meaningless. Maybe not meaningless, I understand it, but it doesn't evoke anything in me."; "I think it's good for such a topic to have such really simple visualisations, that I found quite helpful than just reading the text."
Rating of the decision aid overall	"It leaves a good impression."; "I think it's great to have this kind of information beforehand, I think it's great when you're dealing with this painful topic of mammography."; "The presentation is totally, very unattractive ... most women are not academics. And I think you just have to take a shorter and crisper approach."
Rating of the text	"It all sounds pretty clear."; "The other stuff was all very nice and easy to read and understand."; "Nope, the text is great."

Appendix B. Results of Qualitative Interviews with Themes, Sub-Categories, and Example Quotes (continued)

Themes and Sub-Categories	Example Quotes
Interviewer effects	
Effect of gender	"No, no, I have no problems going to a male gynaecologist, so I would answer you straight there too."; "No, I don't think so. But that's also because we know each other, and I like you. Maybe if it was a complete stranger, although with doctors you also feel comfortable somehow."
Being observed	"Now that may be because you're there, but you're not just anyone ..."; "Anyway, I think it's nice to go through it together."; "You can't say anything, right?"; "Now I do not know exactly what you want from me, not personally, do not misunderstand me."; "I now have read through it very carefully, because I'm doing this with you now."
Other issues	Screening age too late, Technical/medical language, Invitation process and screening location unclear