Original Article

MDM Medical Decision Making

Medical Decision Making
2020, Vol. 40(2) 212–221
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X20904357
journals.sagepub.com/home/mdm
SAGE

When Does the Incremental Risk Format Aid Informed Medical Decisions? The Role of Learning, Feedback, and Number of Treatment Options

Kevin E. Tiede, Felicia Ripke, Nicole Degen, and Wolfgang Gaissmaier

Background. Informed medical decisions require understanding the benefits and risks of treatments. This entails comparing treatment outcomes to a control group. The *incremental risk format* has been recommended as it directly visualizes the differences between treatment and control group in 1 graph, whereas they have to be calculated from 2 separate graphs in the *total risk format*. We investigated when the incremental risk format aids understanding. **Methods.** In 2 experiments, participants received information about medical treatments, either as incremental or total risk format. We assessed *verbatim knowledge* (precise quantitative knowledge), *gist knowledge* (knowledge of essential meaning), and evaluations of the formats. Study 1 ($N = 99$) consisted of only 1 trial with medical information and also assessed recall. Study 2 ($N = 222$) assessed learning across multiple trials and also varied the presence of feedback and the number of treatment options. **Results.** In study 1, the incremental risk format (v. total risk format) led to worse knowledge, recall, and evaluations. In study 2, participants learned to understand the incremental risk format over time, resulting in comparable verbatim knowledge and evaluations as in the total risk format, as well as in even better gist knowledge. Feedback and number of treatment options did not moderate the effect of risk format. **Limitations.** The studies were conducted with nonpatient samples, and study 2 employed hypothetical treatments. **Conclusions.** The incremental risk format was initially less understandable than the total risk format. After a short learning period, however, the incremental risk format resulted in better gist knowledge and was comparable otherwise, which suggests that participants had to get used to that format. This has important implications for the study of new formats.

Keywords
decision aids, icon array, incremental risks, medical decision making, risk communication

Date received: July 13, 2019; accepted: January 14, 2019

Shared decision making is being increasingly promoted to involve patients in medical decisions.[1–4] To make informed decisions regarding medical treatments, patients need to acquire a quantitative understanding of their benefits and risks.[5] However, many people struggle with grasping risk information.[6–8] Therefore, decision aids have been developed to improve medical decision making.[4] These often comprise visual aids such as icon arrays,[9–11] which have been shown to increase the general and precise understanding of risks.[5,12–16]

To fully grasp the benefits and risks of treatments, it is important to compare the treatment outcomes to outcomes that occur in a control (placebo) group. While patients might erroneously think that treatment outcomes can be fully attributed to the treatment, it is

Corresponding Author
Kevin E. Tiede, Social Psychology and Decision Sciences, University of Konstanz, P.O. Box 43, 78457 Konstanz, Germany (kevin.tiede@uni-konstanz.de).

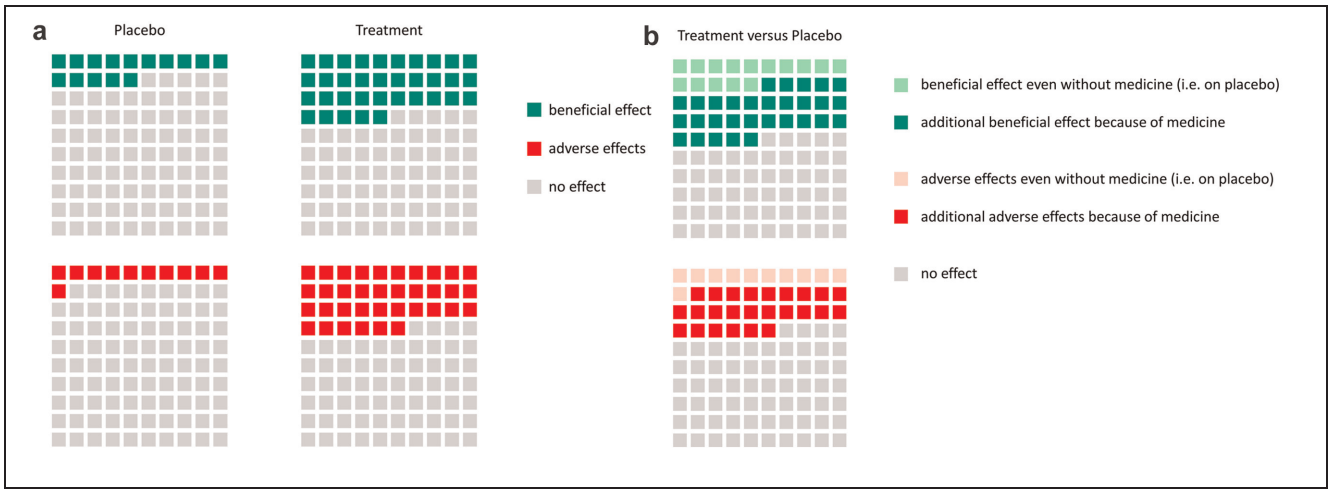

Figure 1 Benefits and risks associated with an exemplary treatment presented (a) in the total risk format and (b) in the incremental risk format.

important that undergoing no treatment or taking a placebo can be associated with these outcomes as well. For example, all people have a certain chance of feeling nauseous, while a medication causing nausea as a side effect only increases this risk beyond the baseline chance.[2,17] To understand the causal impact of the treatment, the crucial information is *how much more frequently* beneficial as well as harmful outcomes occur under treatment compared to under placebo or nontreatment.[16] Commonly, this medical information is presented in icon arrays, which present the information for the treatment and control group separately (*total risk format;* Figure 1a). To assess the causal treatment effect, the total risk format thus requires that people calculate the difference between treatment and control with regard to beneficial and harmful outcomes. The *incremental risk format* (Figure 1b) has been suggested to facilitate this assessment by highlighting exactly the differences between treatment and control in 1 icon array so that people can read those off directly rather than having to calculate them.[16,18]

So far, there exists only little research comparing the total risk format and the incremental risk format. It has been shown that side effects of treatments are subjectively perceived as less common, less worrisome, and less likely when presented in the incremental risk format, compared to the total risk format.[16,18] The reason for this effect presumably is that the incremental risk format helps people avoid erroneously interpreting the entire risk as causal treatment effect, while often there is a certain risk of experiencing the respective side effect (e.g., nausea) without taking a medication. Although the incremental risk format has been argued to counteract this effect,[2,17] so far this argument has not been tested empirically. Furthermore, the incremental risk format has been shown to prevent the order of presented side effects from biasing risk perceptions.[18] Because of these effects, the incremental risk format has been recommended[2,16,17] and used in decision aids.[19] However, with regard to knowledge about benefits and risks, no advantage of the incremental risk format has been found.[16,20]

In sum, the existing evidence supporting the incremental risk format is limited at best. However, we do believe that it is a promising format especially in light of an increasing number of available treatment options that require more comparisons of treatment effects. We therefore conducted 2 studies that investigated whether the incremental risk

Department of Psychology, University of Konstanz, Konstanz, Germany (KET, FR, ND, WG); Graduate School of Decision Sciences, University of Konstanz, Konstanz, Germany (KET); Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany (WG). The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Wolfgang Gaissmaier has received honoraria for lecturing/consultation from Amgen, Bayer, Biogen, Celgene, Genzyme, Merck Serono, MSD, Mundipharma, Novartis Pharma, Sanofi Pasteur, Takeda, and Teva, as well as research support for another study from Biogen. Kevin E. Tiede, Felicia Ripke, and Nicole Degen declare that there is no conflict of interest. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided by the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG) under grant FOR 2374 (DFG Research Unit), the German Research Foundation Centre of Excellence 2117 "Centre for the Advanced Study of Collective Behaviour", and the Graduate School of Decision Sciences at the University of Konstanz. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

format improves understanding of medical information compared to the total risk format and under which conditions this is the case. In line with previous research,[5,15,21,22] we focused on 2 different types of knowledge: *verbatim knowledge* (i.e., precise quantitative knowledge of risks and benefits) and *gist knowledge*[23] (i.e., knowledge of essential meaning). In study 1, we also assessed recall of the information. Besides knowledge, it is important that people perceive the information as comprehensible to increase their willingness to work with the information. Therefore, we also assessed the subjective accessibility of the information and the attractiveness of the representation.

Our studies focus specifically on understanding the boundary conditions under which different formats work well or not, which has largely been neglected in previous research. To foreshadow some of our findings, the results of study 1 suggest that the incremental risk format may initially not be as good as the total risk format with regard to knowledge but may even be worse. Importantly, however, the incremental risk format could have been at a disadvantage, simply because people were not familiar with this relatively novel format. Therefore, study 2 investigated whether a short learning period would be sufficient to at least counteract this presumed disadvantage. To foreshadow more results, this was actually the case, and the incremental risk format was even advantageous for gist knowledge.

As further boundary conditions, we studied whether the success of the incremental risk format depends on the number of treatment options. While the incremental risk format highlights the frequency differences between treatment and control, the total risk format requires people to compute and memorize these differences. Because the differences can be compared directly with the incremental risk format, we expected that the benefit of this format compared to the total risk format will be higher the more treatments have to be compared. Finally, we explored the effect of providing feedback after the knowledge questions (or not), because it was initially not clear how well people learned to use the incremental risk format and whether this learning actually required feedback.

To summarize: the current studies investigated how people understand information on benefits and risks of treatments depending on the presentation format (total v. incremental risk format). Besides knowledge, we also measured subjective ratings of accessibility and attractiveness. In study 1, a pilot study, we explored the 2 risk formats to examine their performance with regard to knowledge, recall, and subjective ratings. In study 2, we investigated the effect of learning across trials, the number of treatment options, and the presence of feedback. We assessed numeracy and graph literacy as covariates in both studies, because both have been shown to affect the comprehension of risk information.[8,16,21,24,25] Both studies were approved by the ethics committee of the University of Konstanz.

Study 1

The first study was designed as an exploratory pilot study to gain insights into how the incremental risk format affects different types of knowledge, recall, and subjective ratings compared to the total risk format. It served as basis for a more detailed investigation of the incremental risk format in the subsequent study.

Methods

Design. Participants were provided with information on the benefits and side effects of 3 painkilling medications (aspirin, ibuprofen, and paracetamol [acetaminophen]). The medical information was identical as and can be found in Gaissmaier et al.,[21] who distilled it from 3 Cochrane reviews. The information was presented using icon arrays displaying graphically how many people out of 10 people experienced the benefit or the side effect with the medication or the respective placebo (i.e., icon array with 10 icons). While the information was equal for all participants, the representation differed between conditions. In the total risk format condition, the information for the treatment and placebo group was displayed in separate icon arrays. In the incremental risk format condition, this information was summarized in 1 icon array. In both conditions, there were separate icon arrays for benefits and side effects (see Figure 1 for an illustration of the 2 formats and Supplement B for screenshots from study 1). In sum, the independent variable was risk format, and the dependent measures were verbatim and gist knowledge, recall, accessibility of the information, and attractiveness of the representation.

Procedure. In the first part of the study (T1) and after providing informed consent, participants were randomly assigned to 1 of the 2 conditions and were provided with the information on the 3 painkillers in the respective risk format. Below the information, participants were asked the verbatim and the gist knowledge questions with 1 question block per page. Next, participants rated subjective accessibility of the information and subjective attractiveness of the representation. Subsequently, numeracy and graph literacy were assessed. In the second part of the study 2 days later (T2), recall was measured by

asking the same verbatim and gist knowledge questions as in T1 but without presenting the medical information again.

Measures. All knowledge questions as well as subjective rating items were previously used by Gaissmaier et al.[21] and can be found in Supplements D to F.

Verbatim knowledge. We assessed verbatim knowledge by using 8 numerical questions, divided into 2 blocks with 4 questions each. The first block asked participants to read off information from the icon arrays (e.g., "If 10 people take ibuprofen, how many of them will experience side effects?"). In the second block, participants should calculate absolute differences between the frequencies of the treatment and the placebo group (e.g., "If 10 people take paracetamol, how many of them will experience side effects, which they would not have experienced by the respective placebo?"). The verbatim knowledge score represents the proportion of correct answers.

Gist knowledge. We assessed gist knowledge with 5 questions in 2 blocks, which asked for ordinal, nonnumerical comparisons of the medications. Whereas the 3 questions of the first block focused on 1 dimension only (i.e., benefits or side effects; e.g., "Which drug caused side effects most frequently?"), the 2 questions of the second block required the integration of both dimensions (e.g., "Which painkiller is best overall?"). The gist knowledge score represents the proportion of correct answers.

Accessibility. We measured the subjective accessibility of the information with 5 questions with a 5-point scale ranging from 1 (*not at all*) to 5 (*very much*). The questions asked to rate the following aspects of accessibility: comprehensibility, usefulness, seriousness, intuitive accessibility, and difficulty to answer questions (reversed item). The accessibility score represents the mean rating of all 5 questions. The internal consistency of the scale was good ($\alpha = .80$).

Attractiveness. We measured the attractiveness of the representation with 8 questions with a 5-point scale ranging from 1 (*not at all attractive*) to 5 (*very attractive*). In particular, we asked to rate the attractiveness of the following aspects: overall impression, colors, imagery, technical implementation, size, font size, font size, font, and composition. The attractiveness score represents the mean rating of all 8 questions. The internal consistency of the scale was very good ($\alpha = .89$).

Numeracy. We assessed numeracy using the adaptive version of the Berlin Numeracy Test (BNT).[26] The BNT is a widely used test of objective numeracy with good psychometric properties.[26] In the adaptive version, it has 2 or 3 numerical questions (depending on performance) and places participants into 1 of 4 numeracy levels.

Graph literacy. We measured graph literacy using the Short Graph Literacy Scale.[27] It contains 4 items that assess comprehension of health-related information presented in graphical formats. The graph literacy score represents the sum of correct answers.

Participants. Participants were recruited via Amazon Mechanical Turk (MTurk) and received $2.00 for participation. Participants had to be located in the United States and had to be successfully approved in at least 95% of their previous tasks. One hundred participants took part in the first part of the study (T1). One participant who completed the study too fast to have responded conscientiously was excluded from analysis. From the remaining sample ($N = 99$), 58.6% were female, 27.3% were 30 years old or younger, and 44.4% had a college degree or more. Of these participants, 69 participants completed the second part of the study. Demographics as well as numeracy and graph literacy did not differ between participants who completed both parts and those who dropped out (see Supplement A for details). Across all 99 participants, mean numeracy was 2.20 (SD = 1.20), and mean graph literacy was 2.37 (SD = 1.04).

Data analysis. To analyze the effect of risk format on knowledge, we conducted a mixed-design analysis of variance (ANOVA) with risk format (Format; total v. incremental) as between-subjects factor and knowledge type (verbatim v. gist) and time (T1 v. T2) as within-subjects factors. Numeracy and graph literacy were included as continuous covariates. Only participants who completed both parts of the study ($N = 69$) were included in this analysis. To analyze the effect of format on accessibility and attractiveness, we conducted 2 ANOVAs with the predictor risk format and the covariates numeracy and graph literacy. Because these measures were only included in T1, these analyses were conducted using all participants who completed T1. The results were similar when the same analyses were conducted without numeracy and graph literacy as covariates.

Results

Verbatim and gist knowledge. Accuracy for both verbatim and gist knowledge was higher with the total risk format (mean = .78, SD = .26) than with the incremental risk format (mean = .65, SD = .32; $F_{\text{Format}}(1, 65) = 18.69$, $P < 0.001$), and this effect did not depend on knowledge type ($F_{\text{Format} \times \text{KnowledgeType}}(1, 67) = 2.29$,

$P = 0.135$). Overall, knowledge scores were higher in T1 compared to recall in T2 ($F_{\text{Time}}(1, 67) = 172.25$, $P < 0.001$). This decline did not depend on the format in which the information was presented ($F_{\text{Time} \times \text{Format}}(1, 67) = 1.93$, $P = 0.170$). This was also true when taking into account knowledge type ($F_{\text{Time} \times \text{Format} \times \text{KnowledgeType}}(1, 67) = 2.54$, $P = 0.116$). Finally, people higher in numeracy and graph literacy achieved higher knowledge scores ($F_{\text{Numeracy}}(1, 65) = 8.92$, $P = 0.004$; $F_{\text{GraphLiteracy}}(1, 65) = 14.31$, $P < 0.001$).

Accessibility and attractiveness. Information was rated as more accessible in the total risk format than in the incremental risk format ($F_{\text{Format}}(1, 95) = 20.64$, $P < 0.001$). Similarly, the total risk format was rated as more attractive than the incremental risk format ($F_{\text{Format}}(1, 95) = 4.79$, $P = 0.031$). There were no effects of numeracy or graph literacy on accessibility or attractiveness ratings.

Discussion

In this study, we compared the incremental risk format with the total risk format and focused on knowledge, recall, and subjective ratings. Surprisingly, the incremental risk format led to worse knowledge and recall than the total risk format. Consistently, the incremental risk format was rated as less accessible and less attractive than the total risk format.

As the incremental risk format is less common, people could have struggled to understand the risk information in this unknown format. Therefore, people may need to learn to get used to the incremental risk format to exploit its full potential. Furthermore, the comparison of medications was fairly easy. In more complex judgment situations with more treatment options, however, the advantage of the incremental risk format (i.e., making the computation unnecessary) could be larger. Therefore, we conducted a second study, in which we investigated the role of learning, the impact of the number of treatment options, and the effect of feedback on knowledge and subjective ratings.

Study 2

The goal of study 2 was to study if people are able to learn to work with the incremental risk format and how task complexity (here: number of treatment options) affects its performance. Specifically, we were interested if the incremental risk format leads to comparable or even better knowledge and subjective ratings than the total risk format when people have the chance of getting used to it. Therefore, we designed a study with 8 subsequent trials of treatment comparisons.

Furthermore, by highlighting the frequency differences between treatment and placebo, the incremental risk format makes computing this information unnecessary. If the environment becomes more complex with an increasing number of treatment options, the incremental risk format could play out this benefit and become more helpful than the total risk format. Therefore, we also manipulated task complexity by varying the number of treatment options. Finally, we were interested in how feedback affects learning and manipulated if participants received feedback on the accuracy of their answers.

Methods

Design. Participants answered questions about information on the benefits and side effects of multiple medications. For 8 diseases, we created hypothetical medications (labeled medication A, medication B, etc.) with respective frequencies on the benefits and side effects of the treatment and the placebo group. We measured knowledge scores for each set of medications. Each information was presented in an icon array with 100 icons, with separate icon arrays for benefits and side effects (Figure 1 and Supplement C).

We manipulated 3 between-subjects factors in a random fashion. First, the medical information was presented in the total risk format or in the incremental risk format. Second, the number of treatment options was manipulated by providing either 3 or 6 medications per disease. Third, half of the participants were given feedback on the accuracy of their answers, while the other half did not receive any feedback.

In total, we implemented a 2 (format; between) $\times$ 2 (3 v. 6 medications; between) $\times$ 2 (feedback v. no feedback; between) $\times$ 8 (trials; within) design for knowledge and a 2 $\times$ 2 $\times$ 2 $\times$ 2 (assessment after first v. last trial; within) design for subjective ratings. While these factors served as independent variables, verbatim and gist knowledge, accessibility, and attractiveness were dependent variables.

Procedure. Only people without colorblindness could participate in the study. Following the random assignment of participants to 1 of the 8 conditions, informed consent, and the instructions, participants were provided with a brief explanation of a disease and information on the frequency of benefits and side effects of either 3 or 6 hypothetical medications for this disease. This information was presented in either the total risk format or the

incremental risk format. Participants had to answer 8 knowledge questions based on this information, 1 question at a time, and the information was visible throughout. In total, there were 8 such trials consisting of 8 different diseases, which were presented in random order. After each trial, participants in the feedback condition were provided with feedback on the accuracy of their answers and, in case of a wrong answer, the correct solution. Participants were asked to rate the accessibility and attractiveness after the first and the last trial only. Numeracy and graph literacy were assessed at the end.

Measures

Verbatim and gist knowledge. We assessed verbatim knowledge with questions similar to those in study 1, except for 2 differences. First, because of the study design with multiple trials, we reduced the number of knowledge questions per trial. Second, we adapted the wording of the questions slightly to the used stimuli. In this study, each block consisted of 2 questions. Whereas the questions of the first block again asked to read off information (e.g., "If 100 people take medication B, how many of them will get better altogether?"), the 2 questions of the second block required to determine the absolute difference between the frequencies of the treatment and the placebo group (e.g., "If 100 people take medication A, how many more of them will get better compared to the respective placebo?"). The verbatim knowledge score represents the proportion of correct answers.

We assessed gist knowledge with 4 questions similar to those in study 1 but also with 2 questions per block and modified wording. The first 2 questions focused on 1 dimension only (e.g., "For which medication was experiencing side effects most frequently observed altogether?"), whereas the 2 questions of the second block required the integration of the 2 dimensions (i.e., benefits and side effects; e.g., "Which medication is best overall?"). The gist knowledge score represents the proportion of correct answers.

Accessibility and attractiveness. We used the same accessibility and attractiveness measures as in study 1. The internal consistency of the accessibility scale was satisfactory ($\alpha = .74$ and .78 for first and second measurement, respectively). The internal consistency of the attractiveness scale was very good ($\alpha = .91$ and .92 for first and second measurement, respectively).

Numeracy and graph literacy. In study 2, we used the nonadaptive version of the BNT so that participants received all 4 possible numeracy questions of the adaptive version. The numeracy score represents the number of correct responses. We used the same graph literacy scale as in study 1.

Participants. Participants were recruited via Amazon MTurk and received a flat fee of $4.00 and a performance-contingent bonus of on average $0.90 (SD = $0.26) for participation. Based on a power analysis considering only between-subjects factors to determine the lower bound for the sample size using G*Power[28] ($\alpha = .05$, $1 - \beta = .95$, $f = .25$), our study required 27 participants per condition (i.e., 216 in total). Anticipating partial exclusion of participants, we collected data of 241 participants. Due to errors in the survey program, the data of 3 participants had to be discarded. Fifteen participants were excluded from analysis because they failed the attention check, and 1 participant was excluded because he or she completed the study too fast to have responded conscientiously. From the remaining sample ($N = 222$), 39.6% were female, the mean age was 35.4 years (SD = 10.7 years), and 46.2% had at least a bachelor's degree. Mean numeracy was 1.55 (SD = 1.40), and mean graph literacy was 2.23 (SD = 1.07).

Data analysis. The data were analyzed conducting a mixed-design ANOVA for knowledge as dependent variable. Between-subjects factors were risk format (total v. incremental), number of medications (3 v. 6) and feedback (yes v. no), and within-subjects factors were type of knowledge (verbatim v. gist) and trial (first to eighth). Numeracy and graph literacy were included as covariates. For accessibility and attractiveness, 2 mixed-design ANOVAs were conducted with risk format, number of medications, and feedback as between-subjects factors and time (after first v. last trial) as within-subjects factor. Numeracy and graph literacy were included as covariates. The analyses showed similar results when conducted without numeracy and graph literacy as covariates.

Results

Verbatim and gist knowledge. Figure 2 illustrates verbatim and gist knowledge scores. Complete results of the full mixed-design ANOVA model can be found in Supplement F. Knowledge was not generally affected by risk format ($F_{\text{Format}}(1, 212) = 0.34$, $P = 0.563$). Risk format did, however, affect verbatim and gist knowledge differently, with the incremental risk format benefiting gist knowledge ($F_{\text{Format} \times \text{KnowledgeType}}(1, 214) = 17.05$, $P < 0.001$).

We proposed that the incremental risk format could be particularly helpful if people learn to get used to it. In general, people were able to learn how to use the risk formats, indicated by improving knowledge scores across trials ($F_{\text{Trial}}(1, 212) = 17.69$, $P < 0.001$). Especially verbatim

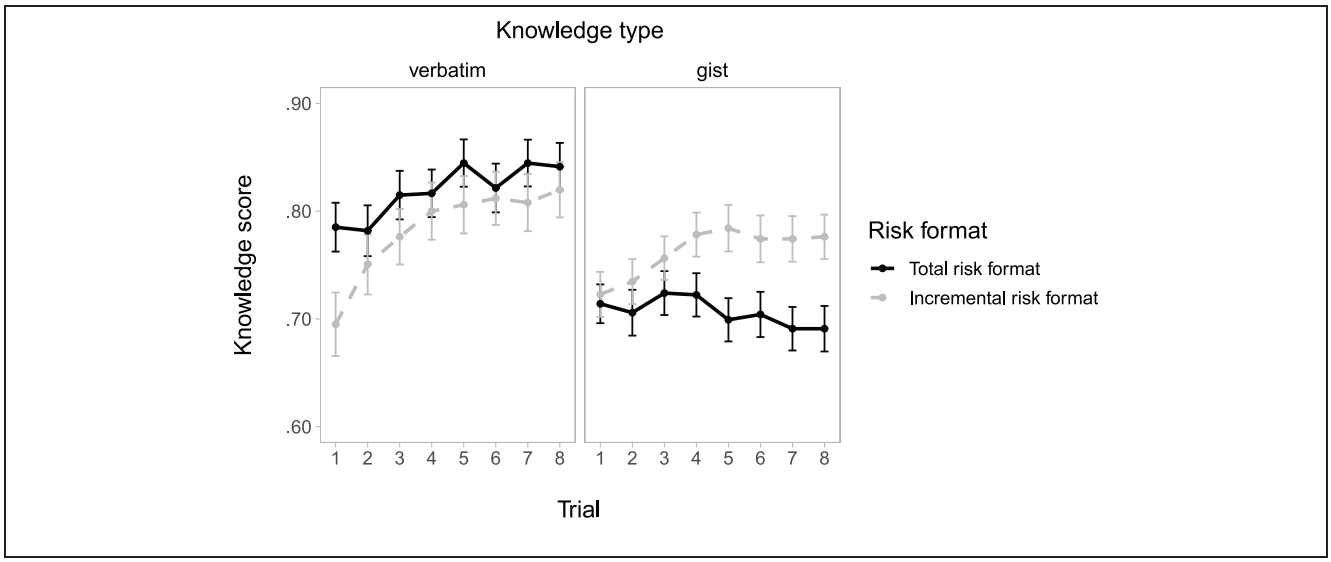

Figure 2 Study 2: while there was no general benefit of either risk format, the incremental risk format was beneficial only for gist knowledge and led to more knowledge across the experiment. Error bars represent 1 standard error of the mean.

knowledge benefited from learning ($F_{\text{KnowledgeType} \times \text{Trial}}(1, 214) = 16.86$, $P < 0.001$). As hypothesized, knowledge increased more strongly with the incremental risk format compared to the total risk format in the course of the experiment ($F_{\text{Format} \times \text{Trial}}(1, 214) = 7.76$, $P = 0.006$).

We further expected that the incremental risk format is especially helpful with higher task complexity. As expected, a higher number of medications as an operationalization of higher complexity led to lower knowledge scores ($F_{\text{Complexity}}(1, 212) = 13.04$, $P < 0.001$). Contrary to our hypothesis, however, format did not affect knowledge differently in less and more complex situations ($F_{\text{Format} \times \text{Complexity}}(1, 214) = 0.20$, $P = 0.658$).

We further examined the effect of feedback on knowledge. Participants who received feedback after each trial generally achieved higher knowledge scores than those who did not ($F_{\text{Feedback}}(1, 212) = 15.19$, $P < 0.001$), which was similar in both risk formats ($F_{\text{Feedback} \times \text{Format}}(1, 214) = 2.30$, $P = 0.131$). Finally, both higher numeracy and graph literacy were associated with higher knowledge scores ($F_{\text{Numeracy}}(1, 212) = 47.67$, $P < 0.001$; $F_{\text{GraphLiteracy}}(1, 212) = 16.13$, $P < 0.001$).

Accessibility and attractiveness. Subjective accessibility and attractiveness ratings are illustrated in Figure 3. Overall accessibility and attractiveness ratings were similar for both formats (accessibility: $F_{\text{Format}}(1, 212) = 0.60$, $P = 0.440$; attractiveness: $F_{\text{Format}}(1, 212) = 0.89$, $P = 0.346$). We proposed that getting used to the risk format would especially benefit accessibility and attractiveness of the incremental risk format. Consistent with this proposal, both accessibility and attractiveness received higher ratings in the total risk format at T1, but this difference disappeared completely at T2 (accessibility: $F_{\text{Format} \times \text{Time}}(1, 214) = 11.65$, $P < 0.001$; attractiveness: $F_{\text{Format} \times \text{Time}}(1, 214) = 9.94$, $P = 0.002$).

Discussion

Study 2 examined how risk format affects knowledge and subjective ratings as a function of learning, number of treatment options, and feedback. The incremental risk format led to lower verbatim knowledge and similar gist knowledge scores than the total risk format before learning (i.e., on trial 1). As hypothesized, however, the incremental risk format benefited more strongly from a short learning period that allowed participants to get used to it than the total risk format. After learning, the incremental risk format even outperformed the total risk format with respect to gist knowledge, and it matched the total risk format with respect to verbatim knowledge. This is in line with a study that found that a practice exercise improved understanding of medical information presented as survival curves.[29] Similarly, accessibility and attractiveness ratings were lower for the incremental than for the total risk format after the first trial but not after the participants had the chance to get used to the format. Knowledge scores were generally higher with fewer treatment options and with feedback, but these

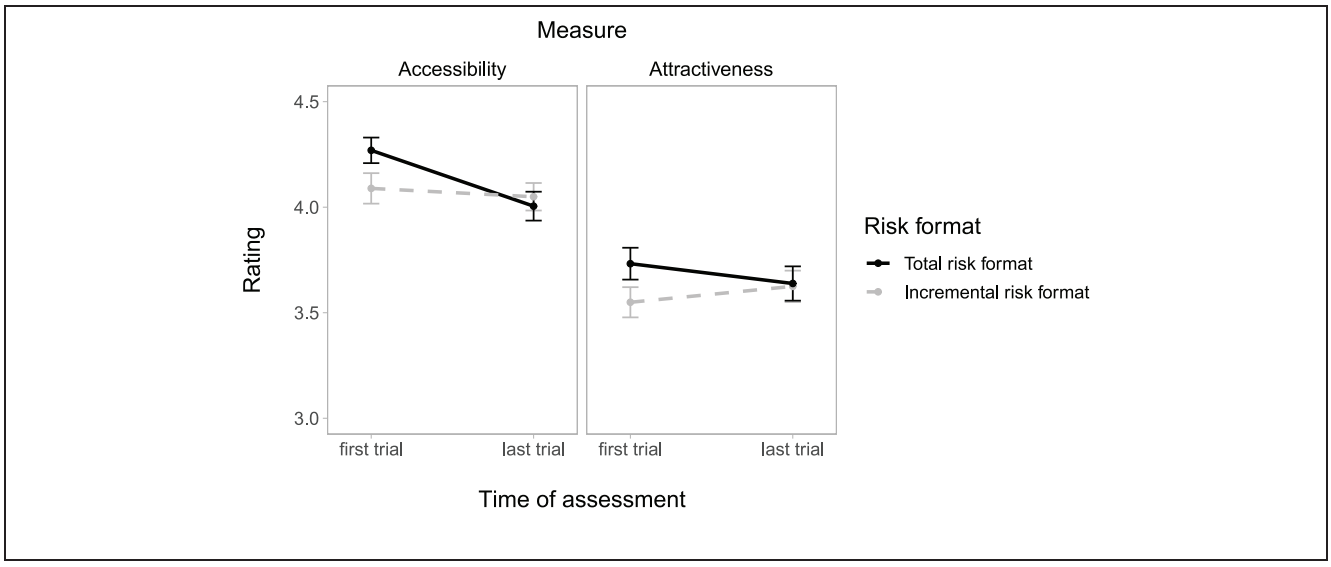

Figure 3 Study 2: after repeated exposure, the incremental risk format (v. total risk format) was rated as better or less bad. Error bars represent 1 standard error of the mean.

effects did not differ between risk formats. In sum, these findings replicate the inferiority of the incremental risk format after a single trial of study 1 for verbatim knowledge and subjective ratings but show that learning makes this inferiority disappear, while it turns the initial equivalence with regard to gist knowledge into superiority.

General Discussion

To facilitate the understanding of incremental benefits and risks, the incremental risk format has been introduced as a variation of an icon array.[16,18] Because there was limited empirical evidence for the benefits of this format, this study aimed to test if and under which conditions the incremental risk format can be better than the more common total risk format.

In 2 studies, participants received information about the benefits and risks of different treatments in 1 of the 2 formats and were asked to answer knowledge questions about them as well as to rate accessibility and attractiveness. Across both studies, the incremental risk format led to lower knowledge scores (except for gist knowledge in study 2) and lower subjective ratings after a single trial, which did not provide the chance of learning. However, study 2 revealed that after learning, the inferiority of the incremental risk format with respect to verbatim knowledge, accessibility, and attractiveness disappeared and the equivalence with respect to gist knowledge turned into superiority. That is, when participants had the chance to get used to the incremental risk format, they were able to exploit its advantages. Fewer (v. more) treatment options and the availability of feedback (v. no feedback) led to better performance overall, but these factors did not moderate the effect of risk format on knowledge.

Limitations

Limitations of our studies have to be considered. First, our studies were not conducted with real patients who actually have to decide about their own treatment. Because they could respond differently, further research is needed to examine the incremental risk format in a patient sample.

Second, study 2 presented hypothetical rather than real treatments to participants. Actual data on benefits and risks of treatments could have led to deviating results. However, the aim of our study was to investigate the general performance of the incremental risk format, so we used medical data with a variety of possible benefits and risks frequencies. This approach is also supported by replicating central results from study 1 (which included real data from Cochrane reviews) in the first trial of study 2.

Third, using different denominators across studies could limit comparability of the results of the 2 studies. The size of the denominator affects knowledge,[16] and a qualitative study found that a denominator of 10 poses the risk of people thinking that only 10 people participated in the medical study.[30] Nevertheless, we do believe that our main conclusions hold true independent of

denominator size, because the results were similar for study 1 and the first trial of study 2.

Implications

Our findings have implications for risk-communication research as well as practitioners. Most importantly, our studies show that the incremental risk format is a promising format to convey benefits and risks of medications and that it is particularly helpful to convey gist knowledge. This is likely useful in many situations, as gist knowledge may often have a larger impact on medical decision making than verbatim knowledge.[5,31,32]

As a caveat, however, it is important to qualify that recommendation and limit it to situations in which patients have the chance to get accustomed to it—at least until the incremental risk format is widespread and common to patients. More research is needed to understand the efficacy of the incremental risk format more comprehensively. For that, it would be helpful to study in more detail features of how such graphs are designed, including denominator size, color, and whether graphs are annotated by numbers. In addition, it would be important to go beyond knowledge and study actual medical decisions, ideally in real patients.

Furthermore, our findings have implications beyond the use of a particular risk format. They encourage to incorporate learning and features of the environment as factors moderating the effect of presentation formats in future research on risk communication. The majority of research on risk formats assesses outcome variables after 1 or a few trials and/or mainly manipulates features of the format or the information.[5,15,16,18,20] Considering learning and the judgment ecology poses a promising approach to better understand the effect of presentation formats and to improve medical decision making. If innovative risk formats such as the incremental risk format have advantages compared to more conventional ones, it should also be considered to educate the public and to use them more often and consistently so that patients are able to work with them.

Conclusion

The objective of our 2 studies was to provide further insights into how to design icon arrays to improve medical decision making. We investigated the promising incremental risk format, with a particular focus on the boundary conditions under which it may be recommendable. Based on our findings, we support the promotion of the incremental risk format[2,16,17] but acknowledge that there are some constraints to it. More broadly, our research emphasizes that learning to use a presentation format has to be considered. Especially when novel formats are compared to more conventional formats, novel formats may be at a disadvantage simply because they are unknown. So far, research on risk communication has largely neglected the conditions under which medical information is processed. We believe our research highlights that varying features of the ecology in the study of risk communication is a fruitful avenue that can yield a more profound understanding how, under which conditions, and for what goals specific presentation formats work well.

ORCID iDs

Kevin E. Tiede https://orcid.org/0000-0002-8735-0630
Wolfgang Gaissmaier https://orcid.org/0000-0001-6273-178X

References

1. Gigerenzer G, Gray JAM. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. Cambridge, MA: MIT Press; 2011.
2. Lin GA, Fagerlin A. Shared decision making: state of the science. *Circ Cardiovasc Qual Outcomes*. 2014;7:328–34.
3. Sheridan SL, Harris RP, Woolf SH. Shared decision making about screening and chemoprevention. *Am J Prev Med*. 2004;26:56–66.
4. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2017;4:CD001431.
5. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Educ Couns*. 2008;73:448–55.
6. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Publ Int*. 2007;8:53–96.
7. Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affair*. 2007;26:741–8.
8. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull*. 2009;135:943–73.
9. Garcia-Retamero R, Cokely ET. Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. *Hum Factors*. 2017; 59:582–627.
10. Spiegelhalter D. Risk and uncertainty communication. *Annu Rev Stat Appl*. 2017;4:31–60.
11. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med Inform Decis Mak*. 2013;13:S7.

12. Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychol*. 2009;28(2):210–16.
13. Garcia-Retamero R, Cokely ET, Wicki B, Joeris A. Improving risk literacy in surgeons. *Patient Educ Couns*. 2016;99: 1156–61.
14. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks. *Soc Sci Med*. 2010;70:1019–25.
15. Tait AR, Voepel-Lewis T, Zikmund-Fisher BJ, Fagerlin A. Presenting research risks and benefits to parents: does format matter? *Anesth Analg*. 2010;111:718–23.
16. Zikmund-Fisher BJ, Ubel PA, Smith DM, et al. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Educ Couns*. 2008;73:209–14.
17. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst*. 2011;103:1436–43.
18. Zikmund-Fisher BJ, Fagerlin A, Roberts TR, Derry HA, Ubel PA. Alternate methods of framing information about medication side effects: incremental risk versus total risk of occurrence. *J Health Commun*. 2008;13:107–24.
19. National Institute for Health and Care Excellence. Statins to reduce the risk of CHD and stroke: patient decision aid. 2014. Available from: https://www.nice.org.uk/guidance/cg181/resources/patient-decision-aid-pdf-243780159.
20. Price M, Cameron R, Butow P. Communicating risk information: the influence of graphical display format on quantitative information perception—accuracy, comprehension and preferences. *Patient Educ Couns*. 2007;69:121–8.
21. Gaissmaier W, Wegwarth O, Skopec D, Müller A-S, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. *Health Psychol*. 2012;31:286–96.
22. Stone ER, Gabard AR, Groves AE, Lipkus IM. Effects of numerical versus foreground-only icon displays on understanding of risk magnitudes. *J Health Commun*. 2015;20: 1230–41.
23. Reyna VF. A theory of medical decision making and health: fuzzy-trace theory. *Med Decis Making*. 2008;28:829–33.
24. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. *Med Decis Making*. 2010;31:444–57.
25. Peters E, Hart PS, Fraenkel L. Informing patients: The influence of numeracy, framing, and format of side effect information on risk perceptions. *Med Decis Making*. 2011; 31(3):432–6.
26. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: the Berlin Numeracy Test. *Judgm Decis Mak*. 2012;7(1):25–47.
27. Okan Y, Janssen E, Galesic M, Waters EA. Using the short graph literacy scale to predict precursors of health behavior change. *Med Decis Making*. 2019;39(3):183–95.
28. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149–60.
29. Armstrong K, FitzGerald G, Schwartz JS, Ubel PA. Using survival curve comparisons to inform patient decision making. *J Gen Intern Med*. 2001;16:482–5.
30. Schapira MM, Nattinger AB, McHorney CA. Frequency or probability? A qualitative study of risk communication formats used in health care. *Med Decis Making*. 2001;21(6):459–67.
31. Reyna VF, Nelson WL, Han PK, Pignone MP. Decision making and cancer. *Am Psychol*. 2015;70(2):105–18.
32. Blalock SJ, Reyna VF. Using fuzzy-trace theory to understand and improve health judgments, decisions, and behaviors: a literature review. *Health Psychol*. 2016;35(8):781–92.