Article:
Bonner, C, Trevena, LJ, Gaissmaier, W et al. (6 more authors) (2021) Current best practices for presenting probabilities in patient decision aids: Fundamental principles. Medical Decision Making. ISSN 0272-989X

https://doi.org/10.1177/0272989X21996328

Current best practice for presenting probabilities in patient decision aids: Fundamental Principles

Running head: Presenting probabilities in patient decision aids

Authors: Carissa Bonner (PhD, carissa.bonner@sydney.edu.au)[1-2], Lyndal J. Trevena (PhD, lyndal.trevena@sydney.edu.au)[1-2], Wolfgang Gaissmaier (PhD, gaissmaier@uni-konstanz.de)[3], Paul K.J. Han (PhD, hanp@mmc.org)[4-5], Yasmina Okan (PhD, y.okan@leeds.ac.uk)[6], Elissa Ozanne (PhD, elissa.ozanne@hsc.utah.edu)[7], Ellen Peters (PhD, ellenpet@uoregon.edu)[8], Daniëlle Timmermans (PhD, drm.timmermans@amsterdamumc.nl)[9], Brian J. Zikmund-Fisher (PhD, bzikmund@umich.edu)[10]

Affiliations:

1. Faculty of Medicine and Health, School of Public Health, The University of Sydney, Australia
2. ASK-GP NHMRC Centre of Research Excellence, The University of Sydney, Australia
3. University of Konstanz, Germany
4. Center for Outcomes Research and Evaluation, Maine Medical Center Research Institute
5. School of Medicine, Tufts University, USA
6. Centre for Decision Research, University of Leeds, UK
7. University of Utah, USA
8. University of Oregon, USA
9. Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands
10. University of Michigan, USA

Corresponding author: Carissa Bonner

Rm 128A, Edward Ford Building A27, The University of Sydney, NSW Australia

Word count:

Abstract = 248 words

Main paper = 5,644 words

No. of Tables = 1

No. of Figures = 0

Funding Statement:

This work was completed with no external funding. The authors received their usual salaries from the organisations outlined under affiliations.

Keywords: decision aids, standards, risk communication

ABSTRACT

Background: Shared decision making requires evidence to be conveyed to the patient in a way they can easily understand and compare. Patient decision aids facilitate this process. This paper reviews current evidence for how to present numerical probabilities within patient decision aids.

Methods: Following the 2013 review method, we assembled a group of 9 international experts on risk communication across Australia, Germany, the Netherlands, UK, and USA. We expanded the topics covered in the first review to reflect emerging areas of research. Groups of 2-3 authors reviewed relevant literature based on their expertise and wrote each section before review by the full authorship team.

Results: Of 10 topics identified, we present 5 fundamental issues in this paper. While some topics resulted in clear guidance (presenting the chance an event will occur, addressing numerical skills), other topics (context/evaluative labels, conveying uncertainty, risk over time) continue to have evolving knowledge bases. We recommend presenting numbers over a set time period with a clear denominator, using consistent formats between outcomes and interventions to enable unbiased comparisons, and interpreting the numbers for the reader to meet the needs of varying numeracy.

Discussion: Understanding how different numerical formats can bias risk perception will help decision aid developers communicate risks in a balanced, comprehensible manner, and avoid accidental 'nudging' towards a particular option. Decisions between probability formats need to consider the available evidence and user skills. The review may be useful for other areas of science communication where unbiased presentation of probabilities is important.

INTRODUCTION

Shared decision making requires evidence-based probabilities to be conveyed to the patient in a way they can easily understand, in order to compare options and come to an informed health decision. This includes the chance of experiencing an adverse health outcome if no action is taken, as well as the chance of experiencing benefits and harms from different medical interventions [1, 2].

Patient decision aids are designed to present quantitative evidence about options in an unbiased way to facilitate shared decision making [1]. The most recent 2017 Cochrane review of patient decision aids includes 105 studies involving 31,043 people and over 50 different decisions [3]. This update continues to provide high quality evidence that decision aids can improve patients' knowledge, and help them feel more informed and clear about their values.

In order to present probabilities in a way that facilitates an informed choice, we need to better understand how people perceive and process risk information. We can then develop decision aids that optimize informed decision-making and account for the heuristics that can influence our perceptions [4]. This understanding may help decision aid developers avoid nudging patients towards a particular intervention or outcome through differences in the way they are presented [5, 6]. Varying skill levels may influence the best format to use for a particular target audience, as certain formats may enable better understanding in patients with low numeracy and literacy [7-10].

We aimed to update a previous review of current evidence for how to present probabilities in a way that maximises understanding, with a focus on informing decision aid development [2].

METHODS

Following the method of the first review published in 2013 [2], the International Patient Decision Aids Standards (IPDAS) collaboration convened international groups of experts on IPDAS quality dimensions, including risk communication (led by BZ and LT). This included researchers from Australia, Germany, the Netherlands, UK, and USA. Several opportunities were also provided for the chapter leads to liaise with other chapter lead authors about overlap and gaps between the standards.

We reviewed and expanded the topics covered in the first review to reflect emerging areas of research. Ten topics were agreed as covering the key aspects of communicating numeric information in patient decision aids. Five of these were considered to be fundamental principles and are detailed in this first paper. The remainder apply numeric information in more complex contexts and are outlined in our second paper. In addition, our discussions led to a new introductory section considering the rationale for (and potential downsides of) numerical presentations of probability. Two to three authors reviewed and wrote each section based on their expertise before bringing the topics together for the full IPDAS report, which was subsequently split into two papers due to the large number of topics covered. Each author took a leadership role on two topics, and all contributed to reviewing the final paper. We resolved disagreements through discussion to reach consensus on which topics were considered fundamental versus advanced. The final fundamental topics included: 1) presenting the chance that an event will occur, 2) putting numerical estimates in context and evaluative labels, 3) conveying uncertainty, 4) time-based risk formats, and 5) skills for understanding numerical estimates. The advanced paper covers: 1) effect sizes of treatment/screening options, 2) personalised risk estimates, 3) visual formats, 4) graph literacy, and 5) interactive/web-based formats. We have continued to provide this update as a non-systematic expert review of the literature on these 10 topics, drawing on the depth and breadth of the

authors' expertise. Following the development of the expert-written section updates, all co-authors further contributed to the update which includes more than 230 references.

Within each section, we have summarised the key points that continue to hold from the previous IPDAS update in 2013 [2], as well as new evidence, with recommendations highlighted. More detail is provided for topics where advancement has occurred since the previous update; with brief overviews for other issues. Where emerging issues for future research and/or future systematic reviews were identified, these are highlighted in the Results and expanded further in the Discussion section.

RESULTS

A preparatory question: Are numbers needed?

A difficult issue for decision aid developers is how much information to include. Including detailed risk and benefit information supports the ethics of informed choice and can reduce over- or under-estimation of risk, which is otherwise both common and substantial [11-18] particularly among less numerate individuals [19]. Numbers may, for some people at least, evoke greater trust [20], and people report wanting quantitative information. As a result, current standards for decision aids support provision of quantitative information as a quality indicator [2].

Yet, interpreting quantitative risk and benefit information can be challenging. As discussed in the below sections, less numerate individuals find the information retrieval and comparison tasks necessary to derive meaning from risk data more challenging. Furthermore, in unfamiliar situations, all patients can

find it difficult to determine whether a risk statistic should be interpreted as good or bad absent use-relevant reference standards or evaluative labels [21, 22].

The value derived from providing quantitative information about risks and benefits appears to depend on patients' ability to derive appropriate gist (including affective) representations from this information [23]. If patients can do so consistently, accurately, and in task-appropriate ways, then the argument for presenting numbers in risk communications is strong. If they cannot, then the question becomes whether non-numerical risk presentations might be more consistently and accurately interpreted [24]. By this argument, less numerical precision (including use of non-quantitative formats such as evaluative category information only) may be appropriate for certain tasks, such as providing clear motivational or behavioral signals (e.g., being at "high risk" for developing a disease implies a need to act), although evidence is needed for this conjecture.

Strong arguments exist for ensuring availability of quantitative risk information for patients. Whether quantitative information about probabilities is always necessary as the initial, primary communication of risk, however, is less clear, and only a few decision aid studies have begun to investigate this question [24]. The data that does exist is difficult to interpret, with studies finding that quantitative information can have different effects in the screening context: increased colorectal screening, no effect on colorectal cancer screening uptake, and decreased cervical cancer screening intentions [25-27]. More generally, there is a need for research evaluating both how targeted quantitative risk information can best be used support informed decision making, and when and under what circumstances tradeoffs may exist regarding use of number-heavy risk presentations.

Evidence and recommendations regarding numerical presentations of probability

What is most clear from our review is that many numerical format and presentation decisions can influence the effectiveness of numerical presentations of probability. The key recommendations from each section are summarised in Table 1, with more detail in text below.

[insert table 1]

Presenting the chance an event will occur

The research evidence regarding optimal methods for presenting the chance that a single event will occur has not substantively changed in the last decade. For both written and verbal information, patients have a more accurate understanding of risk when probabilistic information is presented as numbers rather than words, even though some may prefer receiving words because it seems easier to understand, as one systematic review has shown [28]. Some newer studies show that combining verbal and numerical risk formats may lead to overestimation [29]; therefore, care should be taken if formats are combined.

Suitable formats for presenting numeric chances depend on the nature of the task. When the task is to present the chance of a single event, either frequency formats that express the expected proportion of affected and unaffected individuals within a given time frame (e.g., "Every year, 10 in 100 people with pre-diabetes develop diabetes"), or percentage formats (e.g., "Every year, 10% of people with pre-diabetes will develop diabetes") are more transparent and informative than formats such as "The annual chance of developing diabetes is 10%" [30]. The last statement is problematic because it does not specify the reference class—that is, the population for whom the risk estimate applies (e.g., people with pre-diabetes"). Without a clear description of the reference class, people might draw inappropriate inferences about their own risks, or misinterpret probabilities as referring to event frequencies in their

own lives [30]. For example, a patient taking fluoxetine for mild depression might hear from her doctor that there is a "30-50% chance of developing a sexual problem such as impotence or loss of sexual interest," and misinterpret this statement as indicating that she will have problems in 30% of her own sexual encounters. In other words, she might mistakenly interpret the reference class or denominator of the risk estimate as the total of her own sexual encounters, rather than the total of all patients who take fluoxetine [31].

Presenting risks as frequencies reflects the way that evidence-based objective probability estimates are truly derived. However, there is also some experimental evidence that risks presented as frequencies are perceived as higher than when they are presented in their equivalent percentage value, especially in patients with lower numeracy [32] and (possibly) when smaller percentages are presented [33]. There is some evidence that people find percentages less than one (e.g., 0.1%) more difficult to understand than the equivalent simple frequency (e.g., 1 in 1000) [33]. However, this problem may reflect difficulty manipulating decimal points (e.g., asking someone to represent 1 in 1000 as a percentage) rather than a comprehension problem. Combining simple frequency and percentage formats appears to add no advantage [33].

The use of the "1-in-x" format to communicate probability estimates should be avoided in patient decision aids. Available experimental evidence suggests that "1 in x" formats can elevate probability perceptions leading to biased risk perceptions [34-36] and thereby affect decision making [37, 38]. This effect appears to be small but consistent [36] and does not seem to be more prevalent among people with lower numeracy [36]. Since patient decision aids should be balanced and not designed to persuade toward one particular option, the 1-in-X format is not recommended [39].

When the task is to compare the chance of occurrence of two or more independent events (e.g., the chance of symptom relief with drug A compared with placebo), formats that express the chance of an event using a single number, such as a percentage, work better than frequencies that require the use of more than one number, such as 1 in 100 [33]. If using simple frequencies such as 1 in 100, one should use the *same* denominator (e.g., 1 in 100 versus 2 in 100) as these are easier to compare than frequencies using different denominators (e.g., 1 in 100 versus 1 in 50). Consistent denominators should always be used. Experimental studies have shown that when choosing the size of the denominator, smaller numbers (e.g., 100) are easier to understand and remember than larger numbers (e.g., 10,000) [40]. When more than one probability is communicated, e.g., in situations involving multiple benefits and harms of multiple drugs, organizing the information through the use of a balanced and transparent format is also important. One such format that has been shown to improve consumer understanding and decision-making is the so-called "fact box", a simple table-based summary of benefits and harms [41, 42].

Numerical estimates in context and evaluative labels

As mentioned in the preceding sections, interpreting health-related numerical estimates is an important component of a patient's decision-making process. Traditional approaches to communicating these risks, in fact, have often relied on numerical approaches. However, numeric information in health is often unfamiliar, lacking in meaning, and may be unusable for patients, even when numeric comparisons are available and/or patients believe they are using the numbers [21]. Partly because of this, people often do not comprehend numerical information and can have difficulty using such information in decision making [43]. Some approaches have relied on affective (evaluative) meaning to modulate the communication process [44, 45]. Research suggests that health-related risk estimates can

be provided using numerical information in context, including numeric comparisons and evaluative labels and symbols to improve patient understanding [46].

Directly interpreting the meaning of numeric information (e.g., telling patients how good or bad a 9% risk is) can have a substantial influence on how patients use that information in subsequent decisions. For example, providing evaluative labels for numeric quality-of-care information (e.g., telling decision makers that the numbers represented “poor” or “excellent” quality of care) resulted in greater use of this information in judgments and less reliance on an irrelevant emotional state among the less numerate [21]. Evaluative labels for test results (that a patient’s test was “positive” or “abnormal”) induced larger changes to risk perceptions and behavioural intentions than did numeric results alone [47]. Other evaluative symbols and colour coding also can help people identify high-value options and better understand risk factors (e.g., quality stars, check marks, blue ribbons, coloured dots) [48-51] although placement of the evaluative symbol can matter [49]. The benefits of evaluative coding, however, can be unclear or mixed so that they should be used carefully. The use of evaluative labels, for example, can improve understanding of the general concept of uncertainty (whether it is high or low), but it can lead to value-inconsistent choices [52]. Further research is needed on how such formats affect trust, decision-making processes (e.g., feelings, thoughts), decisions, and behaviours.

In order to make sense of risk information, people may also be provided with comparisons to other risks or to risks of other people (e.g. the average person). According to the evaluability hypothesis [22, 53], providing individuals with information about other risks makes it easier for them to evaluate the magnitude of the risk. One approach is the risk ladder in which people are provided with more familiar risk situations, e.g. the comparative risk of smoking. Research showed that less numerate individuals, in particular, were able to distinguish better between risk levels using this method [54]. However, the

addition of familiar risk scenarios to illustrate very low risks related to prenatal risks of chromosomal anomalies did not help women distinguish better between risk levels [55]. More numerate women, however, did distinguish between risk levels better overall. These conflicting results of comparing risks may be related to the specific comparison risk which was used which may be more familiar to some than to others. The choice of comparison risks has the potential to influence risk perceptions so that this choice needs to be made with careful thought.

Another approach is to compare an individual's health risk with that of other people. In one study, individuals who were told they had a higher than average risk for breast cancer were more inclined to choose risk-reducing treatment than those told they had lower risk, despite the risks being equivalent [56]. This effect may be interpreted as less desirable, as the comparison may lead to a less thorough evaluation of risks and benefits of a medical intervention by individuals. Another way to do this is converting an absolute risk to 'biological age' by comparing to average or ideal risk, discussed below.

Conveying uncertainty

Probability estimates embody and express different types of uncertainty, which may or may not be explicitly communicated to patients. The first type, aleatory or "first-order" uncertainty, pertains to the fundamental indeterminacy or randomness of future events. This idea can apply cleanly to populations, as different outcomes occur to different individuals at known rates even as it remains unknown who experiences which outcome. It fits less well in the context of probability estimates for individual patients, since individuals only experience outcomes, not probabilities. The second major type, *epistemic* or "second-order" uncertainty (also known as "ambiguity"), pertains to a lack of knowledge needed to predict future outcomes, and encompasses uncertainty about the adequacy, reliability, or credibility of probability estimates themselves. The resulting imprecision in probability estimates is then

typically expressed by confidence intervals [57]. Epistemic uncertainty and ambiguity in risk information arise from multiple sources, ranging from limitations in empirical evidence to methodological problems including misspecification of risk prediction models, which limit the precision and accuracy of probability estimates [58].

An understanding of these inherent uncertainties of probability estimates is arguably an essential element of informed decision making [59] . However, the communication of these uncertainties can also be psychologically aversive and needs to be undertaken cautiously, as it may bias choices away from options with ambiguity.

As of the 2013 review [2], only a few studies had explored the optimal methods and outcomes of communicating aleatory or epistemic uncertainty in risk information. The communication of aleatory uncertainty related to probability estimates had been examined in a small number of studies of both textual and novel visual methods of representing randomness, which have suggested that these methods have no significant effect on risk perceptions. However, evidence was lacking regarding their effects on patient understanding and other important outcomes [60-64]. The communication of second-order epistemic uncertainty (ambiguity) had also been examined in only a small number of studies, which have shown mixed effects on outcomes such as patient understanding, perceptions of the credibility of risk information, and medical decision making [57, 65-69].

The communication of uncertainty has attracted increasing attention since the previous review, as attested by the emergence of literature reviews on this topic [70-72]. One framework suggests we can consider the communication of epistemic uncertainty as being 'in*direct*' or '*direct*' [72]. *Indirect* uncertainty refers to the quality of the underlying evidence. *Direct* epistemic uncertainty includes the

uncertainty around numeric estimates and may be depicted as a range, a confidence interval, error bars on a chart, etc. Research on the impact of these different formats show considerable variation in their effects on cognition, affect, trust, behaviour, and decision-making [71-74].

Similarly, little evidence exists on the impact of qualitative strategies for communicating *indirect* epistemic uncertainty regarding numeric risk estimates [75]. Since the inception of the GRADE framework for assessing the quality of evidence, patient decision aids have begun to incorporate the use of qualitative evidence quality ratings that express epistemic uncertainty [76]. Examples include evaluative labels (e.g. 'High' or 'Low') or symbols (e.g. stars, shaded circles etc) to indicate evidence quality. However, there is very limited research on their impact on decision making, [72] and evidence quality is only one type of indirect epistemic uncertainty of relevance to medical decision making.

The communication of epistemic uncertainty has been examined in experimental studies using hypothetical scenarios, which have demonstrated effects consistent with what is often referred to as ambiguity aversion—i.e., diminished preferences for the ambiguous option [77-79]. Recent studies have also examined the extent to which different types of uncertainty are communicated in decision support interventions and clinical practice. Reviews of both existing decision support interventions and web-based cancer prognostic tools have shown that aleatory uncertainty is communicated in a highly variable manner—typically consisting of numeric estimates with little supporting explanation of the meaning of probability—and that epistemic uncertainty is rarely communicated at all [70, 71]. A recent observational study of physician-patient communication in cancer care supports these findings [80].

Time-based risk formats

The previous review [2] highlighted the issue of limited data for long term outcomes given short trial timeframes [81], but identified several ways this might be conveyed to patients when available: a) the chance of a specific outcome at a single point in the future; such as 10-year risk of cardiovascular disease used in estimates of risk/benefit for cholesterol medications; b) chance of an outcome at multiple points in the future, such as presenting the risk of repeat bypass surgery at 5 years and 10 years after the initial procedure; c) mortality or survival graphs showing risks over time, where a balanced approach using both survival and mortality graphs is recommended; d) cumulative future or lifetime chance of an outcome, such as describing cancer risk in patients with BRCA gene mutations; and e) rate of occurrence of an outcome that is likely constant over time, such as the annual risk of pregnancy with a specific birth control method. With such a wide range of ways to present risk over time it is important to ensure outcomes are presented in the same way for different intervention options, to avoid biasing patients towards one over the other. For example, a longer time frame will make the risk appear bigger due to base rate neglect (i.e., ignoring the time period), but a shorter time frame may be more salient and avoid the issue of discounting future outcomes [4]. Several emerging time-related formats with new evidence since the last review are detailed below.

'Biological age' is an increasingly popular risk format that converts an absolute risk to a younger or older age by comparing risk factors to 'ideal' or average levels – resulting in various labels such as real age, fitness age, heart age, lung age, kidney age and bone age [82, 83]. Age is a time-based concept, but biological age could also be conceptualised as an evaluative label or a relative risk, as it involves comparing a person's risk to ideal or average values. There is particular interest in heart age, which is increasingly used around the world [84], but the methods used to calculate these risk formats are highly variable – the same person may get a younger or older heart age depending on the model used and whether they are compared to ideal or average risk factors [85].

In terms of risk communication, heart age formats are often used for “persuasive” behaviour change purposes but they also appear to elicit a greater emotional response, are easier to remember, and increase risk perceptions compared to absolute risk formats. Trials in applied settings have found better risk factor control using heart age interventions compared to usual care (e.g. [86, 87]), but these tended to include multiple behaviour change techniques so the effect of biological age as a risk format cannot be isolated. Direct comparisons between heart age and absolute risk are less conclusive. An experiment to test this [88] found higher emotional responses and risk perception in participants who received a heart age message compared to participants who received a 10-year absolute risk percentage. However, heart age did not increase lifestyle change intentions in this study or another that tested the addition of heart age to 10-year risk [89]. Another experiment found personalised heart age was better recalled than a 5-year risk percentage, but it also reduced credibility, inflated risk perceptions in low risk participants and did not promote behavioural intentions to see a GP or change lifestyle [90]. When hypothetical heart age based on averages was compared to percentage or natural frequency, it improved recall, comprehension, evaluations of the information and led to higher behavioural intention for some but not all outcomes [91]. A rapid review of biological age formats in 2020 concluded that it is unclear whether these formats have a positive impact on lifestyle behaviours overall [82]. Preliminary results from a heart age review [92] presented at SMDM in 2020 concurred that there is limited evidence for heart age formats impacting behaviour but it does affect emotional response and risk perception, so decision aid developers should be mindful of this.

Heart age calculators can work well as a marketing tool to get attention and direct consumers to appropriate clinical assessments or lifestyle change [93], but caution is needed to ensure they are not used to ‘nudge’ low risk people towards medication and tests that they are unlikely to benefit from in

the short term [94]. A registered trial (ACTRN12620000806965) will soon test the effect of adding heart age to a cardiovascular disease prevention decision aid for the first time, to investigate how it affects decision making.

There also appears to be increasing interest in prolongation of life data, which can also be framed as postponement of death [95]. A study comparing absolute risk reduction (e.g. risk reduced by x%) to prolongation of life (e.g. life extended by x months) in relation to statin use found no difference for decision confidence or satisfaction with risk communication, but prolongation of life format reduced prescription uptake [95]. A review including 6 prolongation of life trials found that medication was acceptable if it prolonged life by >8 months for 48% of participants on average, which increased to 64% for 8 months or more [96]. Delay of event formats are analogous to prolongation of life formats if the 'event' is death, but other outcomes such as heart attacks or cancer recurrence may also be presented. One study found that delay of event formats were associated with positive views of treatment benefits, and high willingness to initiate/pay for treatment. There is also some indication that healthcare providers and patients value survival gain differently, for example where doctors' recommendations are associated with their own perceptions of survival gain, and insensitive to patient preferences [97, 98].

Skills for understanding numerical estimates

Numeracy is the ability to understand and apply mathematical concepts. It relates strongly to the use and interpretation of numerical information. Higher numeracy can facilitate computations, the interpretation of numbers, information seeking, depth of processing, and trust in numerical formats, leading to improved risk comparisons, risk estimates, and value elicitations [43, 99]. On the other hand, lower numeracy is associated with overestimation of risk probabilities [100, 101], higher susceptibility to factors other than numerical data (e.g. framing, mood states, labels used to interpret

quantitative results and feedback from others) [21, 102], and more denominator neglect [103, 104]. Higher numeracy is associated with higher education, younger age, being Caucasian, and being male; it differs somewhat across countries [43, 99, 105].

Many people have limited skills for dealing with numeric information in patient decision aids. The OECD estimated that 19% of adults across all 23 OECD countries are at or below Level 1 of numeracy. These least numerate can only do simple numeric operations; they can count, sort, and do basic arithmetic. No such systematic data exist on numeracy in physicians and other health care professionals. The few existing studies are typically based on small convenience samples and suggest that physicians generally have much higher numeracy skills than the general population. However, these data also suggest that there are substantial proportions of physicians that lack numerical skills and/or an understanding of central numerical concepts in medicine (e.g., [8, 9, 30, 106-110]). Even less is known about numeracy in specific patient-disease groups [111].

Numeracy research has focused primarily on 'objective numeracy', that is, the ability to solve numerical problems for which a correct answer exists, or on subjective numeracy as a proxy for objective numeracy [10, 112]. This concept of numeracy has a long tradition beyond the field of making informed medical decisions (e.g., [113]). It became popular with respect to medical decision making through the introduction of a 3-item test [114] which was soon extended [105, 115] [8] [116] [117, 118].

Although useful to have tests that assess objective numeracy, one downside of these tests is that they are intimidating to many people (including physicians, see [119]). A less burdensome alternative is subjective numeracy [10], which assesses the self-perceived ability to handle numerical operations and a general preference for working with numbers. Subjective numeracy predicts comprehension of numeric

information almost as well as objective numeracy [105, 120] but recent research indicates that it is a different construct, capturing more meta-cognitive, emotional, or motivational aspects of decision making [121, 122]. People who lack numeric confidence understand numeric information less well, take fewer actions in decisions that require dealing with numbers, and make worse choices [121, 123, 124]. In fact, the benefits of greater objective numeracy to making good health (and financial) decisions emerge primarily among people with sufficient numeric confidence [125].

Numeracy research has progressed considerably and enough evidence exists in this field now to complete a systematic review although currently only narrative reviews exist [112]. Furthermore, basic research findings in objective numeracy imply that especially less numerate patients and consumers will be helped by decision aid developers using communication techniques that: 1) draw and maintain attention to numeric information; 2) highlight the importance of doing the math for patients; and 3) provide evaluative meaning to numeric information in decision aids [126]. Without additional assistance in decision aids, less numerate people may rely more on compelling stories and emotional reactions in decisions rather than the hard facts and, as a result, potentially may make worse decisions when numbers are involved [45, 127]. Although people may believe they understand and effectively use numbers in their decisions, this is not always the case. Decision aid developers, however, can choose evidence-based methods to present numeric information that will help less numerate patients in particular. Finally, another communication technique that can help less numerate patients is the use of simple visual displays such as icon arrays or bar graphs (see accompanying paper for further details). See also use of evaluative labels above.

DISCUSSION

This review provides key recommendations for decision aid developers deciding how to present numerical probabilities about health outcomes and the risks and benefits of intervention options. The overarching principles include using numerical formats, using consistent risk formats for options and outcomes to make it easy to compare, and testing risk formats with the end users particularly if contextual information like evaluative labels are used or skills are limited. More specifically, we recommend: 1) using frequencies or percentages over a set period of time with a clear denominator to convey probabilities numerically; 2) choosing risk formats that enable unbiased comparisons between outcomes and interventions (e.g. avoid using different 1 in x formats for different options); 3) addressing the needs of patients with varying skill levels by drawing attention to the numbers, doing any mathematical operations for the reader, and using targeted evaluative explanations with care; and 4) using consistent formats across health outcomes and intervention effects wherever possible (e.g. numerical frequency out of 100 over 5 years). The last recommendation is particularly relevant whenever facilitating comparisons and/or considering use of evaluative labels or symbols, representations of uncertainty, or time-based risk formats.

Trade-offs sometimes need to be made between the pros and cons of different probability formats, and such decisions should be made with the aim of presenting options in a consistent and unbiased way. For decision aid developers, the possible risk formats may be limited by the available evidence. For example, if comparing two interventions where outcomes are assessed over 5 years for one but 10 years for the other, we suggest using the 5-year timepoint for both to enable a consistent format that can be compared. Or if comparing two outcomes where one risk is <1% and the other is >1%, we suggest showing this as a frequency out of 1000 people to ensure both risks are presented as a whole number

with the same base rate (instead of requiring decimal points or use of inconsistent base rates). The skills of the intended audience should also be considered when choosing the risk format, and ideally the final materials should be tested with end users in that population.

The review highlights the importance of acknowledging inherent uncertainty when applying objective probabilities of a single event to an individual person. Objective probabilities are empirically observed frequencies of repeated events. However, because an individual has only one life to live - and either will or will not experience the event - frequencies have no logically coherent meaning. At the individual, single-event level, probability estimates represent subjective expressions of confidence, not objective "facts." Decision aids may help this basic level of uncertainty become more transparent to patients, but more complex uncertainty concepts may cause confusion until research identifies more effective communication strategies.

A major limitation of this review is that it was not done systematically, relying instead on the expertise and interests of the authors. We were not able to assess the overall strength of evidence for different risk communication formats using this method, but have highlighted where findings are based on reviews versus single trials. We encourage other researchers to pursue systematic reviews for the numerous topics identified in the paper to inform revisions to future IPDAS guidelines. At least two systematic reviews are currently in progress: 1) comparing the effects of different quantitative risk communication formats across health issues on comprehension, perceptions and decisions (Prospero registration CRD42018086270) [128]; and 2) investigating the quantitative and qualitative effects of communicating CVD risk as heart age on behavioural and psychological outcomes (protocol available as preprint; with preliminary results from 9 quantitative trials presented at SMDM 2020) [92]. Further research is needed in several areas to guide future decision aid developers, including the issue of

conveying different types of uncertainty, how to phrase evaluative labels and comparisons for different cultural groups, and how to choose time points *across* disease areas to avoid biasing perceptions (e.g. if cancer risk is presented over a lifetime while cardiovascular disease risk is presented over shorter timeframes, then cancer will be perceived as a greater risk). Another interesting but relatively unstudied question is whether numerical information is always necessary for every risk communication: are there some contexts where this is unhelpful to patients? Interventions to address the needs of numeracy and health literacy in specific patient sub-groups and amongst health professionals also requires further investigation. This paper may be useful for other areas of science communication where unbiased presentation of probabilities is important, but the overall purpose of communication must be carefully considered – a different approach may be needed if the goal is persuasion, myth-busting or behaviour change.

In conclusion, this paper outlines fundamental principles for decision aid developers to consider when deciding how to present probabilities, updating the earlier IPDAS review [2]. It is complemented by advanced concepts in our accompanying paper. Overall, our recommendation stands to provide numerical risk information to patients, with an emphasis on using consistent formats to improve understanding and unbiased presentation of options in patient decision aids. Understanding how different risk formats can affect risk perception will help decision aid developers avoid accidental 'nudging' towards a particular option.

Table 1: Key recommendations for decision aid developers

Topic	Recommendations
Overarching principles	• In general, use numerical risk formats (versus verbal terms only) for precision, comprehension, and trust building • Consider context and skills when deciding which risk formats to use • Present options and outcomes in the same risk format wherever possible • Test risk formats with end users in the population to whom the risk applies
Presenting the chance an event will occur	• Using frequencies *or* percentages over a specific time period may be considered to reduce information overload • Use consistent denominators and formats across outcomes • Specify the reference class (i.e. the population to whom the risk applies) • It is preferable to avoid 1-in-X formats as they are hard to compare and bias risk perception
Numerical estimates in context and evaluative labels	• Consider using evaluative labels, symbols, or colours to convey gist meaning, but beware of the potential for bias (e.g. different cultural meanings of red) • Consider providing comparative risks and/or reference standards (e.g., in a risk ladder) to clarify meaning, but choice of comparators can influence beliefs and create bias (e.g. anchoring to first number presented)
Conveying uncertainty	• Recognize that people's conceptual understanding of both the 1st-order (related to the fundamental indeterminacy of future outcomes) and 2nd-order (related to the lack of knowledge needed to predict future outcomes) uncertainty embodied by probability estimates is often limited • Be cautious about communicating 2nd-order, epistemic uncertainty (e.g. using probability ranges), given that this uncertainty may be psychologically aversive and difficult to understand, and that optimal methods of communication remain to be determined
Time-based risk formats	• Because little comparative research exists on the multiple ways to convey risk over time (e.g. 5 vs 10 year risk, mortality/survival graphs, cumulative risk, occurrence rate), communicators should consider audience needs when deciding among formats (e.g. what time period is most relevant?) • It is preferable to avoid using biological age to convey 'lifetime risk' because it is not clear to users how this relates to absolute risk or intervention effects, and it can bias risk perception and reduce credibility • Prolongation of life and delay of event information may be useful to help patients weigh up whether the benefit is meaningful to them • If time-based risk formats are used, use the same timeframe for all options and outcomes to avoid bias towards one option
Skills for understanding numerical estimates	• Draw and maintain attention to numeric information • Don't require the reader to do maths; instead, do the math for them, e.g. by calculating and presenting differences or ratios • Help the reader understand a number's evaluative (good or bad) meaning through cues (e.g. labels) and support with visual formats (e.g. icon arrays) • When evaluating the effect of decision aids or risk formats, the numerical skills of both patients and clinicians should be considered

Declaration

The authors have no conflicts of interest to declare.

Acknowledgments

We would like to acknowledge the authors of the previous review on which this paper is based, who were not able to contribute to the update.

References

1. Elwyn, G., et al., *Developing a quality criteria framework for patient decision aids: online international Delphi consensus process.* Bmj, 2006. **333**(7565): p. 417.
2. Trevena, L.J., et al., *Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers.* BMC Med Inform Decis Mak, 2013. **13 Suppl 2**(Suppl 2): p. S7.
3. Stacey, D., et al., *Decision aids for people facing health treatment or screening decisions.* Cochrane Database Syst Rev, 2017. **4**(4): p. Cd001431.
4. Kahneman, D., *Thinking, fast and slow*. 2011, UK: Penguin Random House.
5. Blumenthal-Barby, J.S. and H. Krieger, *Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy.* Med Decis Making, 2015. **35**(4): p. 539-57.
6. Thaler, R. and C. Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness.* 2009, UK: Penguin Random House.
7. Baker, D.W., *The meaning and the measure of health literacy.* J Gen Intern Med, 2006. **21**(8): p. 878-83.
8. Cokely, E.T., et al., *Measuring risk literacy: The Berlin Numeracy Test.* Judgment and Decision Making, 2012. **7**(1): p. 25-47.
9. Garcia-Retamero, R., et al., *Measuring Graph Literacy without a Test: A Brief Subjective Assessment.* Med Decis Making, 2016. **36**(7): p. 854-67.
10. Fagerlin, A., et al., *Measuring numeracy without a math test: development of the Subjective Numeracy Scale.* Med Decis Making, 2007. **27**(5): p. 672-80.
11. Berry, D.C. and M. Hochhauser, *Verbal Labels Can Triple Perceived Risk in Clinical Trials.* Drug Information Journal, 2006. **40**(3): p. 249-258.
12. Berry, D.C., P. Knapp, and D.K. Raynor, *Provision of information about drug side-effects to patients.* Lancet, 2002. **359**(9309): p. 853-4.
13. Berry, D.C., D.K. Raynor, and P. Knapp, *Communicating risk of medication side effects: An empirical evaluation of EU recommended terminology.* Psychology, Health & Medicine, 2003. **8**(3): p. 251-263.
14. Berry, D.C., et al., *Patients' understanding of risk associated with medication use: impact of European Commission guidelines and other risk scales.* Drug Saf, 2003. **26**(1): p. 1-11.
15. Knapp, P., D.K. Raynor, and D.C. Berry, *Comparison of two methods of presenting risk information to patients about the side effects of medicines.* Qual Saf Health Care, 2004. **13**(3): p. 176-80.
16. Lipkus, I.M., *Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations.* Med Decis Making, 2007. **27**(5): p. 696-713.
17. Steiner, M.J., et al., *Understanding risk: a randomized controlled trial of communicating contraceptive effectiveness.* Obstet Gynecol, 2003. **102**(4): p. 709-17.
18. Young, S. and D.M. Oppenheimer, *Effect of communication strategy on personal risk perception and treatment adherence intentions.* Psychol Health Med, 2009. **14**(4): p. 430-42.
19. Peters, E., et al., *Numbers matter to informed patient choices: a randomized design across age and numeracy levels.* Med Decis Making, 2014. **34**(4): p. 430-42.
20. Gurmankin, A.D., J. Baron, and K. Armstrong, *The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication.* Med Decis Making, 2004. **24**(3): p. 265-71.

21. Peters, E., et al., *Bringing meaning to numbers: the impact of evaluative categories on decisions.* J Exp Psychol Appl, 2009. **15**(3): p. 213-27.
22. Zikmund-Fisher, B., *Helping people know whether measurements have good or bad implications: Increasing the evaluability of health and science data communications. .* Policy Insights from the Behavioral and Brain Sciences, 2019. **6**(1): p. 29-37.
23. Zikmund-Fisher, B.J., *The right tool is what they need, not what we have: a taxonomy of appropriate levels of precision in patient risk communication.* Med Care Res Rev, 2013. **70**(1 Suppl): p. 37s-49s.
24. Schwartz, P.H., *Questioning the quantitative imperative: decision aids, prevention, and the ethics of disclosure.* Hastings Cent Rep, 2011. **41**(2): p. 30-9.
25. Hestbech, M., et al., *Effects of numerical information on intention to participate in cervical screening among women offered HPV vaccination: a randomised study.* Scand J Prim Health Care, 2016. **34**(4): p. 401-419.
26. Schwartz, P.H., et al., *Impact of including quantitative information in a decision aid for colorectal cancer screening: A randomized controlled trial.* Patient Educ Couns, 2019. **102**(4): p. 726-734.
27. Schwartz, P.H., et al., *Providing Quantitative Information and a Nudge to Undergo Stool Testing in a Colorectal Cancer Screening Decision Aid: A Randomized Clinical Trial.* Med Decis Making, 2017. **37**(6): p. 688-702.
28. Trevena, L.J., et al., *A systematic review on communicating with patients about evidence.* J Eval Clin Pract, 2006. **12**(1): p. 13-23.
29. Knapp, P., P.H. Gardner, and E. Woolf, *Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations.* Health Expect, 2016. **19**(2): p. 264-74.
30. Gigerenzer, G., et al., *Helping Doctors and Patients Make Sense of Health Statistics.* Psychol Sci Public Interest, 2007. **8**(2): p. 53-96.
31. Gigerenzer, G. and M. Galesic, *Why do single event probabilities confuse patients?* Bmj, 2012. **344**: p. e245.
32. Peters, E., P.S. Hart, and L. Fraenkel, *Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions.* Med Decis Making, 2011. **31**(3): p. 432-6.
33. Woloshin, S. and L.M. Schwartz, *Communicating data about the benefits and harms of treatment: a randomized trial.* Ann Intern Med, 2011. **155**(2): p. 87-96.
34. Oudhoff, J.P. and D.R. Timmermans, *The effect of different graphical and numerical likelihood formats on perception of likelihood and choice.* Med Decis Making, 2015. **35**(4): p. 487-500.
35. Pighin, S., et al., *The 1-in-X effect on the subjective assessment of medical probabilities.* Med Decis Making, 2011. **31**(5): p. 721-9.
36. Sirota, M., M. Juanchich, and J.F. Bonnefon, *"1-in-X" bias: "1-in-X" format causes overestimation of health-related risks.* J Exp Psychol Appl, 2018. **24**(4): p. 431-439.
37. Pighin, S., et al., *Communicating Down syndrome risk according to maternal age: "1-in-X" effect on perceived risk.* Prenat Diagn, 2015. **35**(8): p. 777-82.
38. Sirota, M. and M. Juanchich, *Ratio Format Shapes Health Decisions: The Practical Significance of the "1-in-X" Effect.* Med Decis Making, 2019. **39**(1): p. 32-40.
39. Zikmund-Fisher, B.J., *Continued use of 1-in-X Risk communications is a systemic problem.* Med Decis Making, 2014. **34**(4): p. 412-3.

40. Garcia-Retamero, R. and E.T. Cokely, *Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening.* J Exp Psychol Appl, 2011. **17**(3): p. 270-87.
41. McDowell, M., et al., *Effect of Tabular and Icon Fact Box Formats on Comprehension of Benefits and Harms of Prostate Cancer Screening: A Randomized Trial.* Med Decis Making, 2019. **39**(1): p. 41-56.
42. Schwartz, L.M. and S. Woloshin, *The Drug Facts Box: Improving the communication of prescription drug information.* Proc Natl Acad Sci U S A, 2013. **110 Suppl 3**(Suppl 3): p. 14069-74.
43. Reyna, V.F., et al., *How numeracy influences risk comprehension and medical decision making.* Psychol Bull, 2009. **135**(6): p. 943-73.
44. Brust-Renck, P.G., C.E. Royer, and V.F. Reyna, *Communicating Numerical Risk: Human Factors That Aid Understanding in Health Care.* Rev Hum Factors Ergon, 2013. **8**(1): p. 235-276.
45. Peters, E., *Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. .* Current Directions in Psychological Science, 2012. **21**: p. 31-35.
46. Fagerlin, A., B.J. Zikmund-Fisher, and P.A. Ubel, *Helping patients decide: ten steps to better risk communication.* J Natl Cancer Inst, 2011. **103**(19): p. 1436-43.
47. Zikmund-Fisher, B.J., et al., *Does labeling prenatal screening test results as negative or positive affect a woman's responses?* Am J Obstet Gynecol, 2007. **197**(5): p. 528.e1-6.
48. Damman, O.C., et al., *Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding.* BMJ Qual Saf, 2016. **25**(11): p. 860-869.
49. Greene, J., J.H. Hibbard, and R.M. Sacks, *Summarized Costs, Placement Of Quality Stars, And Other Online Displays Can Help Consumers Select High-Value Health Plans.* Health Aff (Millwood), 2016. **35**(4): p. 671-9.
50. Lazard, A.J., et al., *Website Designs for Communicating About Chemicals in Cigarette Smoke.* Health Commun, 2019. **34**(3): p. 333-342.
51. Oettinger, M.D., et al., *Color-coding improves parental understanding of body mass index charting.* Acad Pediatr, 2009. **9**(5): p. 330-8.
52. Dieckmann, N.F., et al., *Making sense of uncertainty: Advantages and disadvantages of providing an evaluative structure.* Journal of Risk Research, 2012. **15**(7): p. 717-735.
53. Hsee, C., *The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. Organizational Behavior and Human Decision Processes. 1996; 67:247–57.* Organizational Behavior and Human Decision Processes, 1996. **67**: p. 247-57.
54. Keller, C., M. Siegrist, and V. Visschers, *Effect of risk ladder format on risk perception in high- and low-numerate individuals.* Risk Anal, 2009. **29**(9): p. 1255-64.
55. Pighin, S., et al., *Using comparison scenarios to improve prenatal risk communication.* Med Decis Making, 2013. **33**(1): p. 48-58.
56. Fagerlin, A., B.J. Zikmund-Fisher, and P.A. Ubel, *"If I'm better than average, then I'm ok?": Comparative information influences beliefs about risk and benefits.* Patient Educ Couns, 2007. **69**(1-3): p. 140-4.
57. Ellsberg, D., *Risk, ambiguity, and the Savage axioms.* Quarterly Journal of Economics, 1961. **75**: p. 643-669.

58. Spiegelhalter, D. and H. Riesch, *Don't know, can't know: embracing deeper uncertainties when analysing risks.* Philosophical Transactions of the Royal Society A Mathematical, Physical & Engineering Sciences., 2011. **369**(1956): p. 4730-4750.
59. Han, P., *Conceptual, methodological, and ethical problems in communicating uncertainty in clinical evidence. .* Med Care Res Rev, 2013. **70**((1 Suppl)): p. 14S-36S.
60. Ancker, J.S., E.U. Weber, and R. Kukafka, *Effects of game-like interactive graphics on risk perceptions and decisions.* Med Decis Making, 2011. **31**(1): p. 130-42.
61. Baty, B.J., et al., *BRCA1 Testing: Genetic Counseling Protocol Development and Counseling Issues.* J Genet Couns, 1997. **6**(2): p. 223-44.
62. Han, P.K., et al., *Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk.* Patient Educ Couns, 2012. **86**(1): p. 106-13.
63. Lenert, L.A. and D.J. Cher, *Use of meta-analytic results to facilitate shared decision making.* J Am Med Inform Assoc, 1999. **6**(5): p. 412-9.
64. Schapira, M.M., A.B. Nattinger, and C.A. McHorney, *Frequency or probability? A qualitative study of risk communication formats used in health care.* Med Decis Making, 2001. **21**(6): p. 459-67.
65. Han, P.K., et al., *Communication of uncertainty regarding individualized cancer risk estimates: effects and influential factors.* Med Decis Making, 2011. **31**(2): p. 354-66.
66. Han, P.K., et al., *Laypersons' responses to the communication of uncertainty regarding cancer risk estimates.* Med Decis Making, 2009. **29**(3): p. 391-403.
67. Lipkus, I.M., W.M. Klein, and B.K. Rimer, *Communicating breast cancer risks to women using different formats.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(8): p. 895-8.
68. Mazor, K.M., K.S. Dodd, and L. Kunches, *Communicating hospital infection data to the public: a study of consumer responses and preferences.* Am J Med Qual, 2009. **24**(2): p. 108-15.
69. Muscatello, D.J., et al., *Communicating population health statistics through graphs: a randomised controlled trial of graph design interventions.* BMC Med, 2006. **4**: p. 33.
70. Bansback, N., et al., *Communicating Uncertainty in Benefits and Harms: A Review of Patient Decision Support Interventions.* Patient, 2017. **10**(3): p. 311-319.
71. Harrison, M., et al., *Communicating uncertainty in cancer prognosis: A review of web-based prognostic tools.* Patient Educ Couns, 2019. **102**(5): p. 842-849.
72. van der Bles, A.M., et al., *Communicating uncertainty about facts, numbers and science.* R Soc Open Sci, 2019. **6**(5): p. 181870.
73. Dieckmann, N.F., et al., *Seeing What You Want to See: How Imprecise Uncertainty Ranges Enhance Motivated Reasoning.* Risk Anal, 2017. **37**(3): p. 471-486.
74. Dieckmann, N.F., E. Peters, and R. Gregory, *At Home on the Range? Lay Interpretations of Numerical Uncertainty Ranges.* Risk Anal, 2015. **35**(7): p. 1281-95.
75. Han, P., *Conceptual, methodological, and ethical problems in communicating uncertainty in clinical evidence. .* Med Care Res Rev, 2013. **70**: p. 14S-36S.
76. Agoritsas, T., et al., *Decision aids that really promote shared decision making: the pace quickens.* Bmj, 2015. **350**: p. g7624.
77. Bansback, N., M. Harrison, and C. Marra, *Does Introducing Imprecision around Probabilities for Benefit and Harm Influence the Way People Value Treatments?* Med Decis Making, 2016. **36**(4): p. 490-502.

78. Harrison, M., C.A. Marra, and N. Bansback, *Preferences for 'New' Treatments Diminish in the Face of Ambiguity.* Health Econ, 2017. **26**(6): p. 743-752.
79. Visschers, V., *Judgments under Uncertainty: Evaluations of Univocal, Ambiguous and Conflicting Probability Information.* Journal of Risk Research, 2017. **20**: p. 237-255.
80. Engelhardt, E.G., et al., *Disclosing the Uncertainty Associated with Prognostic Estimates in Breast Cancer.* Med Decis Making, 2017. **37**(3): p. 179-192.
81. Lu, C.Y., J. Karnon, and M.J. Sorich, *The importance of high-quality evidence of the long-term impact of nonfatal events used in randomized controlled trials: a case study of prasugrel.* Clin Pharmacol Ther, 2011. **90**(1): p. 27-9.
82. Kulendrarajah, B., A. Grey, and D. Nunan, *How effective are 'age' tools at changing patient behaviour? A rapid review.* BMJ Evid Based Med, 2020. **25**(2): p. 1-2.
83. Spiegelhalter, D., *How old are you, really? Communicating chronic risk through 'effective age' of your body and organs.* BMC Med Inform Decis Mak, 2016. **16**: p. 104.
84. Bonner, C., et al., *Should heart age calculators be used alongside absolute cardiovascular disease risk assessment?* BMC Cardiovasc Disord, 2018. **18**(1): p. 19.
85. Groenewegen, K.A., et al., *Vascular age to determine cardiovascular disease risk: A systematic review of its concepts, definitions, and clinical applications.* Eur J Prev Cardiol, 2016. **23**(3): p. 264-74.
86. Grover, S.A., et al., *Patient knowledge of coronary risk profile improves the effectiveness of dyslipidemia therapy: the CHECK-UP study: a randomized controlled trial.* Arch Intern Med, 2007. **167**(21): p. 2296-303.
87. Lopez-Gonzalez, A.A., et al., *Effectiveness of the Heart Age tool for improving modifiable cardiovascular risk factors in a Southern European population: a randomized trial.* Eur J Prev Cardiol, 2015. **22**(3): p. 389-96.
88. Soureti, A., et al., *Evaluation of a cardiovascular disease risk assessment tool for the promotion of healthier lifestyles.* Eur J Cardiovasc Prev Rehabil, 2010. **17**(5): p. 519-23.
89. Witteman, H.O., et al., *Animated randomness, avatars, movement, and personalization in risk graphics.* J Med Internet Res, 2014. **16**(3): p. e80.
90. Bonner, C., et al., *Is the "Heart Age" Concept Helpful or Harmful Compared to Absolute Cardiovascular Disease Risk? An Experimental Study.* Med Decis Making, 2015. **35**(8): p. 967-78.
91. Damman, O.C., et al., *The effects of infographics and several quantitative versus qualitative formats for cardiovascular disease risk, including heart age, on people's risk understanding.* Patient Educ Couns, 2018. **101**(8): p. 1410-1418.
92. Bonner, C., et al., *Protocol for a systematic review of qualitative and quantitative effects of cardiovascular disease risk communication using 'heart age' concepts.* medRxiv, 2020: p. 2020.05.03.20089938.
93. Bonner, C., et al., *Experiences of a National Web-Based Heart Age Calculator for Cardiovascular Disease Prevention: User Characteristics, Heart Age Results, and Behavior Change Survey.* JMIR, 2020. **22**(8): p. e19028.
94. Bonner, C., et al., *Is the NHS 'Heart Age Test' too much medicine?* Br J Gen Pract, 2019. **69**(688): p. 560-561.
95. Harmsen, C.G., et al., *Communicating risk using absolute risk reduction or prolongation of life formats: cluster-randomised trial in general practice.* Br J Gen Pract, 2014. **64**(621): p. e199-207.

96. Albarqouni, L., J. Doust, and P. Glasziou, *Patient preferences for cardiovascular preventive medication: a systematic review.* Heart, 2017. **103**(20): p. 1578-1586.
97. Halvorsen, P.A., O.G. Aasland, and I.S. Kristiansen, *Decisions on statin therapy by patients' opinions about survival gains: cross sectional survey of general practitioners.* BMC Fam Pract, 2015. **16**: p. 79.
98. Halvorsen, P.A., R. Selmer, and I.S. Kristiansen, *Different ways to describe the benefits of risk-reducing treatments: a randomized trial.* Ann Intern Med, 2007. **146**(12): p. 848-56.
99. Lipkus, I.M. and E. Peters, *Understanding the role of numeracy in health: proposed theoretical framework and practical insights.* Health Educ Behav, 2009. **36**(6): p. 1065-81.
100. Weinstein, N.D., et al., *Colon cancer: risk perceptions and risk communication.* J Health Commun, 2004. **9**(1): p. 53-65.
101. Woloshin, S., et al., *Women's perceptions of breast cancer risk: how you ask matters.* Med Decis Making, 1999. **19**(3): p. 221-9.
102. Peters, E., et al., *Numeracy and decision making.* Psychol Sci, 2006. **17**(5): p. 407-13.
103. Reyna, V. and C. Brained, *Numeracy, ratio bias, and denominator neglect in judgments of risk and probability.* Learning and Individual Differences, 2008. **18**: p. 89-107.
104. Garcia-Retamero, R. and M. Galesic, *Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison.* Am J Public Health, 2009. **99**(12): p. 2196-202.
105. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med, 2010. **170**(5): p. 462-8.
106. Anderson, B.L., et al., *Statistical literacy in obstetricians and gynecologists.* J Healthc Qual, 2014. **36**(1): p. 5-17.
107. Garcia-Retamero, R. and E.T. Cokely, *The Influence of Skills, Message Frame, and Visual Aids on Prevention of Sexually Transmitted Diseases.* Journal of Behavioral Decision Making, 2014. **27**(2): p. 179-189.
108. Hanoch, Y., et al., *Choice, numeracy, and physicians-in-training performance: the case of Medicare Part D.* Health Psychol, 2010. **29**(4): p. 454-9.
109. Jenny, M.A., N. Keller, and G. Gigerenzer, *Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany.* BMJ Open, 2018. **8**(8): p. e020847.
110. Petrova, D.G., J. van der Pligt, and R. Garcia-Retamero, *Feeling the Numbers: On the Interplay Between Risk, Affect, and Numeracy.* Journal of Behavioral Decision Making, 2014. **27**(3): p. 191-199.
111. Gaissmaier, W., et al., *Numeracy of multiple sclerosis patients: A comparison of patients from the PERCEPT study to a German probabilistic sample.* Patient Educ Couns, 2018. **101**(1): p. 74-78.
112. Peters, E., *Innumeracy in the wild: Misunderstanding and misusing numbers.* 2020: Oxford University Press.
113. Paulos, J.A., *Innumeracy: Mathematical illiteracy and its consequences.* 1988: Macmillan.
114. Schwartz, L.M., et al., *The role of numeracy in understanding the benefit of screening mammography.* Ann Intern Med, 1997. **127**(11): p. 966-72.

115. Lipkus, I.M., G. Samsa, and B.K. Rimer, *General performance on a numeracy scale among highly educated samples.* Med Decis Making, 2001. **21**(1): p. 37-44.
116. Weller, J.A., et al., *Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach.* J Behav Decis Mak, 2013. **26**(2): p. 198-212.
117. Schapira, M.M., et al., *The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory.* Med Decis Making, 2012. **32**(6): p. 851-65.
118. Schapira, M.M., et al., *Development and validation of the numeracy understanding in Medicine Instrument short form.* J Health Commun, 2014. **19 Suppl 2**(0 2): p. 240-53.
119. Gaissmaier, W., B.L. Anderson, and J. Schulkin, *How do physicians provide statistical information about antidepressants to hypothetical patients?* Med Decis Making, 2014. **34**(2): p. 206-15.
120. Zikmund-Fisher, B.J., et al., *Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitations.* Med Decis Making, 2007. **27**(5): p. 663-71.
121. Peters, E. and P. Bjalkebring, *Multiple numeric competencies: When a number is not just a number.* J Pers Soc Psychol, 2015. **108**(5): p. 802-22.
122. Waters, E.A., et al., *Examining the Interrelations Among Objective and Subjective Health Literacy and Numeracy and Their Associations with Health Knowledge.* J Gen Intern Med, 2018. **33**(11): p. 1945-1953.
123. Låg, T., et al., *The Role of Numeracy and Intelligence in Health-Risk Estimation and Medical Data Interpretation.* Journal of Behavioral Decision Making, 2014. **27**(2): p. 95-108.
124. Traczyk, J., et al., *Numerate decision makers don't use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making.* Judgment and Decision Making, 2018. **13**: p. 372-381.
125. Peters, E., et al., *Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes.* Proc Natl Acad Sci U S A, 2019. **116**(39): p. 19386-19391.
126. Peters, E., L. Meilleur, and M. Tompkins, *Numeracy and the Affordable Care Act: Opportunities and Challenges. Health Literacy and Numeracy: Workshop Summary. Roundtable on Health Literacy; Board on Population Health and Public Health Practice;* . 2014, Institute of Medicine.: Washington (DC).
127. Bruine de Bruin, W., et al., *Effects of Anti- Versus Pro-Vaccine Narratives on Responses by Recipients Varying in Numeracy: A Cross-sectional Survey-Based Experiment.* Med Decis Making, 2017. **37**(8): p. 860-870.
128. Ancker, J., M. Demetres, and D. Delgado, *Evidence-based communication of numbers in health.* PROSPERO, 2018. **2018 CRD42018086270** (Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018086270).