Taxonomies for synthesizing the evidence on communicating numbers in health: Goals, format, and structure

Jessica S Ancker, MPH, PhD[1]

Natalie C Benda, PhD[2]

Mohit M Sharma, MPH[2]

Stephen B Johnson, PhD[3]

Stephanie Weiner, MS[2]

Brian J Zikmund-Fisher, PhD[4]

1. Vanderbilt University Medical Center, Department of Biomedical Informatics, Nashville, TN
2. Weill Cornell Medicine, Department of Population Health Sciences, New York, NY
3. New York University Langone Health, Department of Population Health, New York, NY
4. University of Michigan School of Public Health, Department of Health Behavior and Health Education, Ann Arbor, MI

Contact Info:
Jessica S Ancker, MPH, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave., Rm 14122
Nashville, TN 37203
jessica.s.ancker@vumc.org
646-248-9281

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/risa.13875.

ABSTRACT

Many people, especially those with low numeracy, are known to have difficulty interpreting and applying quantitative information to health decisions. These difficulties have resulted in a rich body of research about better ways to communicate numbers. Synthesizing this body of research into evidence-based guidance, however, is complicated by inconsistencies in research terminology and researcher goals. In this article, we introduce three taxonomies intended to systematize terminology in the literature, derived from an ongoing systematic literature review. The first taxonomy provides a systematic nomenclature for the outcome measures assessed in the studies, including perceptions, decisions, and actions. The second taxonomy is a nomenclature for the data formats assessed, including numbers (and different formats for numbers) and graphics. The third taxonomy describes the quantitative concepts being conveyed, from the simplest (a single value at a single point in time) to more complex ones (including a risk-benefit trade-off and a trend over time). Finally, we demonstrate how these 3 taxonomies can be used to resolve ambiguities and apparent contradictions in the literature.

KEY WORDS: numeracy, health numeracy, risk communication, data graphics, taxonomy

1. INTRODUCTION

One of the most common questions in health communication is, “What’s the best way to communicate numbers?” This question arises whenever communicators seek to improve understanding of the chances of disease or death; the likelihoods of benefit or harm from vaccines, medical therapies, or lifestyle change; the meaning of laboratory test results, nutrition data, or “quantified self” personal tracking data; or any of the other types of health data that are increasingly available to patients and the public.

Many people, especially those with low numeracy skills, are known to have difficulty interpreting and applying health-related numbers (Ancker & Kaufman, 2007; Lipkus et al., 2001; Peters et al., 2011; Peters et al., 2006; Reyna et al., 2009; Schwartz et al., 1997; B. Zikmund-Fisher et al., 2007). Individuals with low numeracy make suboptimal risk-related decisions in multiple domains(Peters et al., 2006; Schwartz et al., 1997) and are less willing to engage in shared medical decision-making (Galesic & Garcia-Retamero, 2011). Challenges in interpreting numbers can impair patients’ ability to interpret and take appropriate action on the basis of their own medical data, such as the laboratory test results routinely available in electronic patient portals (Zikmund-Fisher et al., 2017). Numbers-related mistakes in interpreting instructions for medications are common among all patients, but especially those with low literacy and numeracy (Bailey et al., 2009; Lokker et al., 2009; Yin et al., 2010).

Although these known difficulties sometimes lead communicators to avoid using numbers altogether (Anderson et al., 2011; Freeman & Bass, 1992; Neuner-Jehle et al., 2011), they have also prompted a rich body of research about better ways to communicate quantitative information. Enough has been published that, for example, the International Patient Decision Aid Standards (IPDAS) Collaboration recently updated its narrative review of methods for presenting probabilities in decision aids.(Bonner et al., 2021; Trevena et al., 2021) As a whole, this literature shows that different ways of communicating quantitative information – variations in number formats or graphic designs – can have important effects on readers.

To provide evidence-based answers to questions about how to effectively present numerical information, we are engaged in a systematic literature review (Prospero registration number CRD42018086270) of the research on communicating probabilities and amounts in health. This broad NIH-funded review includes peer-reviewed experimental or quasi-experimental research that: compares different formats for presenting numbers in any domain of health or medicine, including number formats (e.g., 10% versus 1 in 10 versus 10 in 100) and graphics (such as icon arrays, line graphs, bar charts, and other data graphics); measures quantitative outcomes related to risk perception, comprehension, decisions, preferences, or behavior; and studies samples of lay individuals without medical expertise. Our team has screened more than 36,000 articles retrieved from Medline and 8 other databases to identify 406 papers for inclusion in the review.

However, when we begin to synthesize this literature, we quickly identified apparent inconsistencies. For example, studies by Grimes and Snively (Grimes & Snively, 1999), Siegrist (Siegrist et al., 2008), and Graham (Graham et al., 2009) all recommended against presenting a probability as 1 in X (e.g., 1 in 350). Yet Fair and colleagues (Nagle et al., 2009) and Nagle et al (Nagle et al., 2009) both concluded that the 1 in X format was the best communication option. What explains these seemingly contradictory findings?

As in any analysis of the peer-reviewed literature, variations in the quality and sample size of the studies can sometimes lead to variable findings. However, in examining this literature more closely, it becomes clear that the more serious problem is that different researchers are defining "effective communication" very differently. Grimes and Snively concluded that a communication was effective if a patient could compare 2 numbers and identify the larger of them (Grimes & Snively, 1999). They demonstrated that patients were far more successful at comparing 2.6 and 8.9 per 1000 women than the mathematically equivalent 1 in 384 and 1 in 112 (Grimes & Snively, 1999). By contrast, Fair and colleagues considered a communication effective if it increased a patient's perception of their heart disease risk (Fair et al., 2008). This team (like many other researchers) found that the 1 in X format increased perceived risk compared to other formats such as a percentage, and therefore concluded that it was superior (Fair et al., 2008). Finally, the Graham, Siegrist, and Nagle teams all identified the preferred format as the best communication option. While Nagle found that people preferred the 1 in X format to a percentage (Nagle et al., 2009), Graham (Graham et al., 2009), and Siegrist (Siegrist et al., 2008) found the opposite. This small set of studies suggest that if the goal is to help the reader make an informed choice between 2 risks, then the 1 in X format is less effective than many others. However, if the goal is to increase concern about a health threat, then the 1 in X format is more effective than several alternatives. It also suggests that patient preference is variable and context-dependent; as discussed below, our synthesis suggests that no single number format is universally preferred in all situations.

These examples demonstrate a phenomenon that has become a guiding principle of our work, which is that the literature makes sense only when it is classified by the goal of the communication. In other words, the very idea of a communication method being effective requires answering the question, "Effective at what?"

We have also found that it is not possible to organize this literature or to draw conclusions about the relative efficacy of different communication methods without recognizing that different types of quantitative information carry inherently different cognitive challenges. Interpreting a single laboratory test result is simpler than making a risk-benefit trade-off between 3 different possible therapies for a disease. Different types of data lend themselves to different information formats.

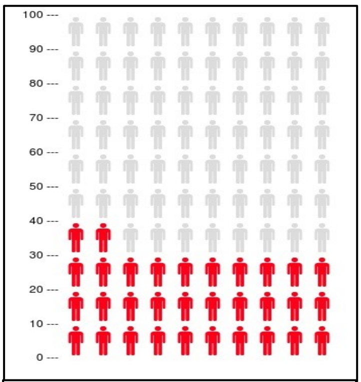


Figure 2: Commonly used risk graphic with at least 28 names

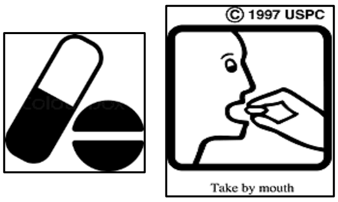
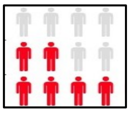


Figure 1: Different visual images can all be called "icons."

If these challenges were not enough to confuse health communicators and researchers alike, a final organizing challenge for this literature is the lack of a uniform vocabulary to describe the ways that we present information. To date, we have discovered 28 synonyms for the graphic depicted in Fig. 1: some of these are *icon array, pictograph, pictogram, figures, crowd chart, crowd figure pictogram, people charts, matrices, pictorial displays, Cates plot*, and *personograph*. Conversely, the terms in the literature are ambiguous. For example, the term *icon* can be applied to one of the component elements of Figure 1 or to the other images in Figure 2.



Similar ambiguity and redundancy problems are found in the terminology for numbers: "*1 in 100*" has been called a frequency, a 1-in-X frequency, a ratio, or simply a probability (and, erroneously, a natural frequency).

To address these issues and synthesize the evidence, we introduce 3 taxonomies developed to assist in organizing the literature on communicating numbers. These taxonomies are likely to require additional development and refinement (for example, by unpacking important subdivisions within categories), but this organizational structure can help clarify degrees of similarity and difference to aid both practitioners and researchers. Furthermore, we provide examples of how using taxonomies of outcome measures, quantitative concepts, and data presentation formats and manipulations can help to resolve ambiguities and apparent contradictions in the literature.

2. TAXONOMY OF OUTCOME MEASURES

In the literature, outcomes are sometimes described with specific terms from behavioral theories (e.g., "perceived susceptibility" from the health belief model(Becker, 1974) or "perceived threat" from the extended parallel process model(Witte, 1992)). But many times they are described using non-specific terms such as "comprehension," "understanding," or "gist understanding."(Reyna, 2008) As described in fuzzy trace theory, gist understanding (a fuzzy representation of the meaning of the information) is often more central to memory and cognition than verbatim understanding (ability to repeat an item of information).(Reyna, 2008) However, in studies of communication effectiveness, researchers must make a judgment about which interpretation represents the most important meaning conveyed by the stimulus, and then select instruments to capture whether the individual has correctly extracted the gist. Because of the lack of clarity associated with these terms, the terms "comprehension" and "gist understanding" are heavily context-dependent in our literature and have hence been measured using a wide range of ad-hoc measures, creating a challenge in comparing and synthesizing the studies. Marteau and colleagues considered that women understood their cervical cancer risk if they considered a "1 in 5000 chance" to be "unlikely" (rather than either "likely" or "impossible").(Marteau et al., 2001) However, respondents in a Zikmund-Fisher et al study demonstrated understanding by identifying how many patients had died at different time points on a survival curve.(B. J. Zikmund-Fisher et al., 2007) By contrast, Pighin et al considered that respondents understood a risk if they could perform computations including doubling a risk and calculating its complement (the chance of *not* getting the condition).(Pighin et al., 2015) Garcia-Retamero and colleagues considered participants to have understood a risk message if they could estimate the relative risk reduction associated with a treatment (for example, a treatment that reduced risk from 8% to 5%).(R. Garcia-Retamero et al., 2011) Although these tasks have all been described generally as understanding or comprehension, they in fact describe cognitively different tasks.

To bring clarity to the literature, in our taxonomy, we avoid the terms comprehension, understanding, or gist. We instead classified studies according to the specific task assessed by the research. For example, in one study, the gist interpretation identified as the correct one might be that the individual *recognizes that the risk is high,* or later *remembers* that she falls into the high-risk category. In other studies, it might be the message that *risk increases with age*, or that a patient's current cholesterol *level is elevated*.

We sought an organizational model that would help us group similar constructs. The Wickens human information processing model (Figure 3) (Wickens et al., 2013) is one of the most widely used models in human factors engineering, where it is routinely applied to explaining and predicting challenges in communication and use of technology, including mistakes in interpretation of written and electronic information. This model describes roughly sequential stages of information processing: *sensory processing*, *perception*, *decision,* and *action*. *Memory* is a resource involved at multiple points in information processing. These distinctions reflected the diversity of outcome measures we noted in the literature review.

2.1: Sensory processing

Sensory processing, during which sensory stimuli enter short-term memory, is a necessary precursor to subsequent information processing. Some studies in the communication domain have assessed sensory processing through variables such as eye fixations (number of times the eye stops scanning to fix on a single location) or gaze dwell time (duration of eye fixation on a single location).(Keller et al., 2014; Kreuzmair et al., 2016; Smerecnik et al., 2010) For example, in a communication about laboratory values and their interpretations, the researcher could measure whether a format change affected how many participants were able to fixate on or visually locate the relevant item of information.

2.2. Perception: Perception covers multiple ways of deriving meaning from the sensory input, both affectively and cognitively.(Becker, 1974; Witte, 1992)

2.2.1 *Affective perceptions*

Many of the studies in our literature review assess **affective perception**, that is, feelings about the risk being communicated, such as worry, concern, or fear. Affective responses may also be collected regarding other types of information, such as laboratory values. Affective risk perceptions are believed to be important predictors of health behavior.(Becker, 1974; Witte, 1992)

2.2.2 *Perceived magnitude*

The perceived size of a risk or a quantity is often of interest. In studies of risk perception, perceived likelihood of a risk may be captured with measures such as "how big does this risk seem to you?" or "how likely is it that you will experience this side effect?" In studies of communication about quantities such as lab values, participants may be asked whether a value seems large or small. Our category of perceived magnitude covers both these types of perceptions. It also does not distinguish between perceived size of a risk and perceived susceptibility to a risk. Although these constructs are separable in theory, in practice it was challenging to determine which was being targeted by an ad hoc measure developed for a particular study. Perceived magnitude of a risk is also typically correlated with affective risk perception, although they may diverge from each other.

2.2.3 *Cognitive perceptions*

Cognitive perceptions assessed in various studies include tasks such as comparing 2 numbers to determine which is larger (**identification of the dominant option**) or stating whether a lab value is elevated, normal, or below normal (**classification**). As described above, this classification focuses on the specific cognitive tasks assessed rather than using a general term such as “gist understanding” that might be difficult to compare across studies. Cognitive perceptions are important because they may mediate actions such as disease management or preventive behaviors as well as monitoring (i.e., affecting whether someone pays attention to future data).(McAndrew et al., 2008)

Some studies have assessed “comprehension” by determining whether participants could perform a computation. For example, Garcia Retamero et al. assessed whether participants who were told a percentage risk could compute the number affected out of 1000 people.(Rocio Garcia-Retamero et al., 2011) Cuite and colleagues assessed whether participants who were given the baseline risk and the relative risk reduction could compute the post-intervention risk.(Cuite et al., 2008) To complete tasks such as these, the participant must draw upon existing knowledge and skills to determine what computation is expected, then identify which numbers in the message are relevant, and perform the computation correctly.(Kirsch, 2001) Task performance is therefore mediated by, and confounded by, numeracy.(Ancker & Kaufman, 2007; Schwartz et al., 1997; B. Zikmund-Fisher et al., 2007) In our taxonomy, this cognitive perception outcome is represented as **performing a computation**.

2.2.4 *Perceptions of the communication*

In addition to assessing perceptions of the risk being communicated, many studies have also captured perceptions of the format in which the information was conveyed. For example, a number of studies have looked at whether patients prefer numbers or graphics in describing risk, or how much they like a specific type of graphic. This type of perception we classified as

engagement with the information. Others assessed **trust in information**, a measure of how much the recipient believes and is willing to rely on the information.

2.3. Decisions/behavioral intentions: Other studies assess decisions, intended decisions, or behavioral intentions, such as whether the respondent chose a particular therapy, expressed the intention to seek more information, or planned to change their behavior. Such **behavioral intentions** are generally accepted as precursors to health behavior (Ajzen, 1991), and are usually simpler to assess because they can be measured at a single point in time, whereas actual behavior of interest must usually be measured over an extended time period.

2.4. Action: Some studies follow respondents all the way to an actual health **behavior**, such as whether participants received cancer screening. We separated actions from decisions/behavioral intentions in light of known gaps between intention and action in health behaviors (e.g., (Rhodes & de Bruijn, 2013)). Although behavior is likely to be a more important outcome measure, it may take hours to years to manifest, and so relatively few studies assessed this outcome.

2.5. Memory: Multiple studies have assessed the *recall* of information in different formats, e.g., assessing whether patients remembered their cholesterol level or cancer risk after receiving it. As described above, fuzzy trace theory distinguishes between verbatim recall of a specific item of information (e.g., a number) and gist recall of a fuzzy representation of meaning.(Reyna, 2008) We classify **verbatim recall** of a specific number or label as a separate outcome, but as described above under "Perception," we sought to disambiguate gist recall tasks by classifying them under the specific cognitive activities measured (including classifying or identifying the normative option). Memory, particularly gist memory, plays a proximal role in the cognitive processing of information and the subsequent formation of intentions.

Table 1: Information processing stage, instruments, and related constructs

Information processing element	Outcome measure category	Outcome measure sub-category	Sample instruments or metrics	Synonyms and closely related constructs
Sensory perception	Sensory perception		Eye fixations or gaze dwell time	Visual attention
Perception	Affective risk perception		How worried are you about this event?	Perceived risk,* perceived susceptibility, perceived threat, worry, concern
	Perceived magnitude		How likely do you think you are to experience this side effect?	Perceived risk*
	Cognitive perception	Performing computation	If your risk now is 5%, what would it be after taking a drug that halves your risk?	Comprehension, understanding, gist understanding
		Identifying dominant option	Which of these values is the highest?	Comprehension, understanding, gist understanding
		Classification	Is this blood sugar level elevated?	Comprehension, understanding, gist understanding
		Ability to detect small changes/sensitivity to deviation	If the risk changed from 10% to 10.5%, would you consider that an increase in risk, or about the same?	Comprehension, understanding, gist understanding
		Ability to identify the direction of a trend over time	Judging by the line graph provided, is the risk increasing, decreasing, or staying the same?	Comprehension, understanding, gist understanding
		Ability to estimate quantity from unlabeled graphic	Estimate what percentage of the people in this graphic are blue	Comprehension, understanding, gist understanding
	Perceptions of the communication	Engagement with information	Which of these graphics do you prefer?	Preference, perceived helpfulness, perceived understandability
		Trust in information	How trustworthy do you find the information provided?	Trust, credibility, believability
Decision	Behavioral intention		Do you plan to get cancer screening?	Response selection, decision

Action	Behavior		Receipt of cancer screening (self-report or objective)	Response execution, behavior, behavior change
Memory	Verbatim recall		What was your blood pressure last week?	Recall

Sensory processing is listed here for completeness but was not a focus of the review.

**Note that both affective risk perception and perceived risk likelihood are often labeled "perceived risk" in the literature.*

.

3. RELATIONSHIPS BETWEEN OUTCOME MEASURES

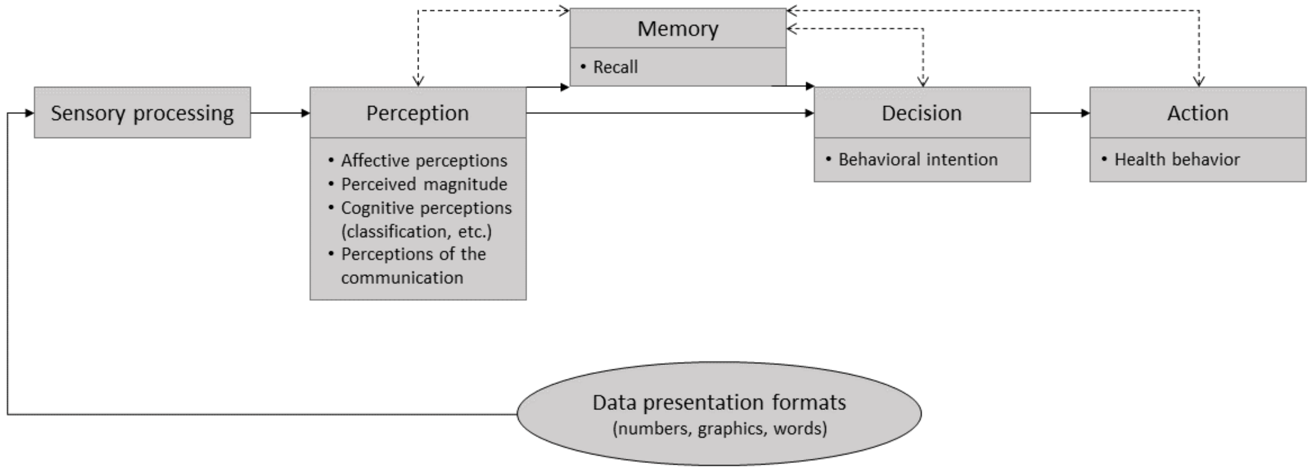
A key lesson from our evidence synthesis is that specific information format may have different effects on the different outcomes. As mentioned above, the evidence is fairly strong that the 1-in-X format for expressing risks leads to higher cognitive and affective risk perceptions than the percentage format. For example, telling someone that their risk is 1 in 1000 is likely to lead to greater concern than saying that the risk is 0.1%. However, the evidence is also strong that for a cognitive task such as comparing two risks to identify the larger, the percentage format is superior. In other words, comparing 1% to 2% is simpler for most people than comparing 1 in 100 and 1 in 50.

Another lesson emerging from this project is that the evidence is much stronger for certain outcomes than for others. For example, among our studies assessing the effects of communications of a single risk, most measure cognitive and affective perceptions such as perceived risk, but only a small number study behavioral intention, and even fewer actual behaviors. As a result, researchers seeking evidence about the relationship between communication and behavior are likely to have to settle for evidence on perceptions or behavioral intentions.

Finally, our synthesis is also suggesting that the evidence around the engagement outcome is weak and contradictory. No single information format, whether it is numbers or graphics, reliably wins popularity contests. We speculate that preferences for informational formats are strongly influenced by the types of formats presented, the familiarity of the formats, and how the respondent intends to use the information. As a result, we suggest that researchers approach this type of research with caution.

The overall message of this literature synthesis is that to derive actionable guidance from this literature, it is critical to identify which of these outcome measures is of interest.

Figure 3: Taxonomy of outcome measures in studies of number communication, with sample metrics



Note: Solid lines link the stages of information processing, while dashed lines indicate cognitive resources that can be accessed at different stages.

4. TAXONOMY OF DATA PRESENTATION FORMATS AND MANIPULATIONS

As described above, a key barrier to synthesizing this literature is the lack of a uniform vocabulary to describe the different manipulations of information tested in the studies. Icon array (Figure 1) has at least 28 synonyms, and conversely, terms such as “icon” and “pictograph” are used to describe very different items. Standardizing format terminology was therefore an important goal of our project.

In our taxonomy, quantitative data can be presented to the reader in 3 basic *data presentation formats: words, numbers,* and *graphics*.

Words: The only words within scope for us were verbal expressions of probability, such as “high risk” or “uncommon,” or of magnitude, such as “large” or “small.” (This meant that we excluded the vast literature on risk messaging through non-quantitative verbal descriptions, such as an explanation of how genes shape heredity in a study of risk of congenital conditions, or images of hazards.)

Numbers: Probabilities can be formatted as percentages (e.g., 10%), proportions (0.10), 1 in X frequencies (e.g., 1 in 10), N in X*N frequencies (e.g., 10 in 100), and odds (e.g., 1 to 9 odds). The relationship between probabilities (for example, a risk increase or decrease) can be formatted as a relative risk using a singular number (e.g., "2x more likely," "half as large"), an arithmetic difference (e.g., "a five percentage point difference," "an absolute risk reduction of 3.5%"), or in rare cases the number needed to treat (NNT). For quantities, number formats are relatively simple, consisting of a number plus unit of measure (for example, 10 mg/dL of cholesterol).

Graphics: We considered only data graphics within scope, that is, graphics showing either probabilities (likelihood of benefit or harm) or quantities (such as radon level, air quality index, or hemoglobin a1c level).(Ancker et al., 2006) A comprehensive list of graphics studied is beyond the scope of this paper, but the most common were number line, bar chart, line graph, and icon array. A *number line* is a 1-dimensional data graphic with a single axis. A *bar chart* is a 2-dimensional graphic with a categorical axis and a continuous axis. A *line graph* is a 2-dimensional graphic with 2 continuous axes. An *icon array* is a graphic depicting the numerator of a probability, the denominator, or both using small icons such as blocks or human figures. We also captured common features of the graphics that can also affect communication. For example, the *part-to-whole property* is applicable to any graphic depicting a probability (for example, a bar chart or an icon array), and describes whether it illustrates the numerator in context of the denominator to show the part in relation to its whole, or the numerator alone.(Ancker et al., 2006) These common graphics also had many variants and subtypes. For example, *survival* and *mortality curves* are cases of line graphs in which the X-axis depicts time and the Y-axis depicts probability.

In addition to classifying the format of the quantitative data, we also classified *format manipulations*, which are commonly studied alterations that can be applied to either numerical or graphical formats.

Framing: One of the most intensively studied format manipulations is *framing,* the choice of whether to depict a number as a gain or as a loss; framing can be applied to either numbers or graphics.(Tversky & Kahneman, 1981) Gain-framing emphasizes the benefit or positive outcome (e.g., a 90% chance of survival), while loss-framing draws attention to the potential harm (e.g., a 10% chance of death).

Context: Another intensively studied format manipulation is the addition of *contextual elements,* which are additional concepts intended to place an otherwise abstract number in context. Examples of commonly studied contextual elements we identified included the use of *verbal interpretive labels* on graphics, the presence or absence of a *population average number* in the communication, the presence or absence of *comparison risks* (e.g., visualization of risks of other diseases in a graphic about colon cancer), and the presence or absence of *anecdotes* or stories about individuals affected by the health condition or risk.

A less-frequently studied set of manipulations was the application of *animation* to graphics (in which visual elements move) and *interactivity* to communications (in which message recipients perform tasks or otherwise manipulate the presentation of the information). Several studies examined the presence or absence of *uncertainty* (e.g., confidence intervals or bands) or contrasted different formats for presenting uncertainty.

In the literature, we see studies of contrasts within and across the 3 major format groups (words, numbers, and graphics), as well as studies of the different format manipulations listed above. The literature can thus be classified into 13 groups of studies (Table 2).

Table 2: Contrasts in the literature between data presentation formats, manipulations, and contextual elements

Contrasts in the literature	Comparators	Example data format (comparator 1)	Example data format (comparator 2)
Contrasts between data presentation formats	Numbers vs. Numbers	The chance is 10%	The chance is 10 in 100
	Words vs. Words	This level is high	This level is above average
	Graphics vs. Graphics	Vertical bar chart with bars portraying 15% and 30%	2 icon arrays of 100 icons each, showing 15 and 30 colored icons respectively
	Numbers vs. Words	The chance of side effects is 0.5%	Side effects are rare
	Numbers vs. Graphics	Your platelet count is 135	Number line display showing 135 on a range from 0 to 500
	Words vs. Graphics	This side effect is common	Icon array showing 10 out of 100 icons as experiencing a side effect
Manipulations of data presentation formats	Framing manipulations	The chance of treatment success is 90%	The chance of treatment failure is 10%
	Presence/absence of or format of animation/Interactivity	Static display (table or graphic)	Animated display that makes numbers visible one at a time
Representation of uncertainty	Presence/absence of or format of uncertainty	The risk is 21%.	The risk is 18% to 24%.
Contextual information added to data presentation formats	Contextual information: Interpretive labels	Your LDL cholesterol is 142 mg/dL	Your LDL cholesterol is 142 mg/dL, which puts you in the borderline high range.
	Contextual information: Population averages	Your risk of getting colon cancer in the next 10 years is 3%	Your risk of getting colon cancer in the next 10 years is 3%. The average person your age has a 10-year risk of 2%.
	Contextual information: Comparison risks	The risk of developing lung cancer is 10 in 1000	The risk of developing lung cancer is 10 in 1000. The risk of developing colon cancer is 1 in 1000.
	Contextual information: Anecdotes	The chance of your chest pain returning is 50%	The chance of your chest pain returning is 50%. For example, David started feeling pain again about 6 months after his procedure.

5. TAXONOMY OF DATA STRUCTURES

The literature is extremely heterogeneous in terms of what types of data are being presented to patients within a single communication (Table 3). As we will show in this section, the communications being studied are constructed from different numbers of populations, variables, and times, so we have called the elements in Table 3 “data structures.” Distinguishing between data structures is important because they pose different cognitive challenges, and because they are communicated with different data presentation formats.

First, data structures involve either probability concepts (such as the chance of harm or the chance of benefit) or quantity concepts (such as radon level, cholesterol level, or patient-reported level of pain).

Both probabilities and quantities can derive from one or more populations. In the simplest situation (left-hand column of Table 3), readers are given information about a single population, i.e., data about themselves or about a defined population. At a higher level of complexity, the communication may present information about multiple populations simultaneously (e.g., average breast cancer rates in several countries). An even more complex situation is the one in which patients are given information about pairs of populations with and without a factor of interest to demonstrate the effect of that factor (right-hand column, Table 3). This category includes data about people with and without exposure to a disease risk factor or a therapy, and it also includes patients before and after receipt of therapy in the many cases when the trend over time is not of particular interest.

Data structures also vary in the number of variables being communicated at once. A relatively simple communication (first row of Table 3) focuses on a single variable such as a patient’s Gail model risk of breast cancer or today’s air quality index. More complex communications may simultaneously present several variables (e.g., the patient’s estimated risks of breast, lung, and colorectal cancers). Variables have *valence*, that is, an affective quality that may be negative (e.g., chance of cancer) or positive (e.g., chance of survival). A special case of multiple variables in a single communication is the trade-off, which is the simultaneous presentation of 2 variables with opposite valence, e.g., the probability of one or more harms and the probability of one or more benefits (trade-offs category of Table 3).

Finally, some data structures highlight time trends (bottom row of Table 3). A survival or mortality curve shows probabilities changing over time, while a line graph of laboratory values shows quantities changing over time.

In general, data structures in the top left of Table 3 are cognitively simpler than those to the right or lower down. For example, interpreting a single number presented in isolation (top row of Table 3) is easier than synthesizing risk and benefit information to make a trade-off decision (central row of Table 2) or making sense of time trends (bottom row). A small number of studies have compared the effects of trends over different time periods (e.g., 5 years vs 10 years).

For simplicity, Table 3 groups together several similar data structures with “one or more” populations or variables. However, in reality, increasing numbers of populations and variables add cognitive complexity, which is depicted more fully in Appendix A. For example, within trade-offs, communications are simpler with a single option and more complex when they present more options. Also, a list, dashboard, or table of numbers representing different populations is more cognitively challenging than information specific to a single patient or population. In general, when the data structure includes data about multiple populations, the structure is not tailored to a primary communication purpose. Instead, the reader must search and identify the relevant pieces of information among other distracting information to answer questions.(Kirsch, 2001)

Sometimes researchers call attention to the difference between simple data structures and more complex ones (e.g., Leonhardt’s study(Leonhardt & Robin Keller, 2018) on communicating a single side effect versus simultaneously communicating the risks of multiple side effects). In many other studies, however, the varying degrees of complexity in the data structure are not discussed, leading to the potential for confounding if studies with more and less complex data structures are grouped together.

Table 3: Taxonomy of data structures, with examples

CATE-GORY	Trend over time	Variable numbers, valences	SINGLES One or more populations[1]	EFFECTS One or more <u>pairs</u> of populations, with and without (or pre- and post-) factor of interest[2]
BASIC	Not of interest	One or more variables, any valence	Probability: *This patient's estimated risk of breast cancer is 4%.* Quantity: *This patient's Ha1c level is 6.5%.* Probability: *Average risk of breast, lung, and colorectal cancer in 5 different countries* Quantity: *Average lipid values in 5 different countries*	Probability: *Treatment reduces cancer recurrence risk from 7% to 3%.* Quantity: *Before therapy, Ha1c level averages 8.0%. After therapy, it averages 6.5%.*[1] Probability: *In younger patients, smoking raises risk by a factor of 4 for lung cancer and 5 for oral cancers. In older patients, [parallel statistics provided]* Quantity: *For men, the therapy would be expected to increase HDL by X mg/dL and reduce LDL by Y mg/dL. For women, [parallel statistics provided]*
TRADEOFFS	Not of interest	Two tradeoff variables (one positive and one negative)[3] for one or more options	Probability: *With the therapy, chance of survival is 80% and chance of blood clot is 0.01%.* *Probability: For men, the chance of survival with Therapy A is 80% and risk of blood clot is 0.01%, while the chance of survival with Therapy B is 90% and risk of blood clot is 0.05%. For women, [parallel statistics provided].*	Probability: *With the therapy, survival chances rise from 70% to 80%, while risk of a blood clot rises from 0% to 0.01%.* *Probability: For men, Therapy A increases survival from 70% to 80% and increases risk of blood clot from 0.01% to 0.02%, while Therapy B increases survival from 70% to 90% and risk of blood clot from 0.01% to 0.05%. For women [parallel statistics provided].*
TIME TRENDS	Of interest	One or more variables, any valence	Probability: *Cancer recurrence risk for a patient for each year over the next 10 years* Quantity: *The patient's systolic blood pressure every month over the past year* Probability: *Risks of lung and colorectal cancers among patients in the US, Canada, and Mexico over a 10-year period*	Probability: *Difference in cancer recurrence risk over the next 10 years between those who do and do not take therapy* Quantity: *The projected effect of medication on systolic blood pressure over 12 months of administration* Probability: *Effect of therapy on survival chances for high, medium, and low risk patients over time* Quantity: *Projected effect of therapy on individual's systolic and diastolic blood pressure over the coming year, for men and women*

1. A population may be a group of any size; a population of size 1 is the patient him- or herself.

2. Pre-post comparisons when the time trend is <u>not</u> of primary interest are depicted in the right-hand column as "pairs of populations." When the time trend between pre-treatment and post-treatment is of primary concern, the communication is classified in the bottom rows as a time trend.

3. Quantity examples are omitted from trade-off cells for space reasons and because they are relatively rare in trade-off messaging.

6. GUIDANCE Our taxonomies of outcomes, data formats, and data structures allow us to classify the literature and answer research questions in this general structure: "For data structure X, what is the effect of data format Y on outcome Z?" This structure helps us to disambiguate and classify the rich literature on this topic in a way that strengthens our ability to synthesize meaningful evidence. For example, if we apply this research question structure to the issues described in the introduction of this paper, we can formulate several relevant questions.

1. "For a comparison of probabilities for 2 populations, what is the impact of the 1 in X format on ability to identify the largest probability?" Grimes and Snively (and other researchers addressing this question) confirm that people are more likely to be able to compare probability numbers with the same denominator (the N in X*N format) than probability numbers with different denominators (the 1 in X format).
2. "For the probability in one group, what is the effect of the 1 in X format on perceived risk?" Fair and colleagues, as well as a number of other researchers, are in agreement that describing a probability in the 1 in X format tends to produce higher risk perceptions than describing it as a percentage.
3. "For a probability in one group, what is the effect of the 1 in X format on patient preference?" The studies summarized earlier, as well as other studies in our literature review, do not suggest that any particular number or graphic is strongly preferred by majorities of patients in different situations. Instead, preference appears to depend on factors such as the patient's familiarity with the format or graphic in question and how they intend to use the information.

Overall, the evidence falls into a matrix (Figure 4).

Data structure	Outcome groups: Affective perception	Perceived magnitude	Cognitive perception	Decision	Action	Recall
One population, one variable, one time, probability						
One population, one variable, one time, quantity						
One pair of populations, one variable, over time, probability						
One pair of populations, one variable, over time, quantity						
One population, two trade-off variables, one time, probability						
Etcetera						

Affective perception

- Evidence on the effects of number formats
- Evidence on the effects of graphic formats
- Evidence on the relative effects of numbers vs graphics
- Evidence on the effects of verbal probabilities
- Evidence on the relative effects of numbers vs verbal
- Evidence on the effects of framing
- Evidence on the effects of showing uncertainty
- Evidence on the effects of animation/interactivity
- Evidence on the effects of context

7. DISCUSSION

Numerical information such as laboratory values and health risks is an essential component of all health communication. The challenge of effectively communicating numerical information has spurred a wealth of research. Making sense of this literature, however, is complicated by lack of standardization of terminologies and organizing principles. We present an organizational structure based on 3 taxonomies. The taxonomy of <u>outcomes</u> organizes the different sorts of outcome measures studied, including sensory processing, perceptions, decisions, actions, and recall. This taxonomy is particularly important, providing a key to unlocking the literature by grouping it by the goal of the communication. It also provides the basis for constructing evidence-based guidance that can be accessed by encouraging researchers and practitioners to select a goal for their communication. The classification of data structures describes the range of types of quantitative concepts being presented, ranging from simple (a single value for a single group) to more complex (e.g., trade-offs and trends over several groups). The classification of data presentation formats and manipulations describes the different ways of presenting each quantitative concept, including different sorts of numbers (e.g., percentages and ratios), graphics (including icon arrays, number lines, and pie charts), and manipulations that can be applied to either numbers or graphics (e.g., framing). Specifying these concepts allows us to use the literature to answer questions in the format, "For data structure X, what is the effect of data format Y on outcome Z?" We can be confident that we are comparing apples to apples.

Our taxonomies have some similarities to categorizations published in previous reviews, but they also have important differences. For example, the 2021 IPDAS reviews of presenting probabilities in decision aids(Bonner et al., 2021; Trevena et al., 2021) mirrored some elements of our data structures taxonomy, in that they separated single-probability communications from effect communications and time-trend communications from single time probabilities. However, the IPDAS papers do not systematically categorize the research by different outcomes (e.g., behavior vs decisions vs recall) to demonstrate that the same format can have different impacts on different outcomes, and sometimes conflate multiple distinct cognitive processes (e.g., performing computations, identifying dominant options). In addition, some parts of the IPDAS papers are structured by data structure (e.g., time trend vs single time), and other parts by data format (e.g., numbers vs graphics). By contrast, our taxonomies permit the systematic categorization of the literature across all 3 dimensions of data structure, data format, and outcome, highlighting the potential for unique effects of each on task-relevant outcomes. Finally, the IPDAS papers were not concerned with the communication of quantities, such as air quality data or laboratory values delivered through patient portals.

There are also similarities between our taxonomies and the 5-factor framework proposed by van der Bles et al. for communicating epistemic uncertainty (van der Bles, 2019). The van der Bles framework focuses on (a) *who* communicates (b) *what*, (c) *in what form*, (d) *to whom*, (e) *to what effect*. Our taxonomy resembles theirs but goes considerably deeper in terms of data structures (corresponding to the Bles category of "what" is communicated), data formats (corresponding to their "in what form"), and outcomes ("to what effect"). However, as demonstrated in Table 2, uncertainty is only one factor of interest in our review. We consider communications in which there is no uncertainty about the numbers, and we do not address uncertainty about non-numerical facts and concepts.

Michie and colleagues have delineated terms for behavior change interventions (Michie et al., 2011), which include communication techniques such as verbal persuasion and persuasive argumentation as well as other techniques such as punishment, classical conditioning, and restructuring the physical environment. Unlike the Michie taxonomy, ours covers communications designed to support outcomes other than behavior change, such as informed decision making (e.g., making trade-off choices between different therapies) and cognitive tasks in support of disease management (such as categorizing one's laboratory result as low, normal, or elevated).

A limitation of our approach is that our literature review included only research on health-related concepts and decisions. Although this nonetheless led to a very large sample of more than 400 eligible papers, we recognize that even more highly relevant literature is found in other domains such as engineering risk management, financial decision-making, and environmental communication. However, because our taxonomies focus on the characteristics and structures of data, data formats, and human information processing, we expect that the central concepts should be transferable to other domains with minimal adaptation. Another potential limitation is that we were unable to classify otherwise relevant studies in which outcomes such as "comprehension" or "understanding" were measured with an ad-hoc battery of knowledge, recall, and computation questions that were aggregated into a single outcome measure (e.g.,(Housten et al., 2020)). Finally, we did not attempt to classify studies that involved multiple interventions concurrently, such as a comparison of a text-only patient information flyer with a novel flyer with reworded text and multiple graphics.

Moving forward, we anticipate that our taxonomies may be useful not only for making sense of the existing literature but also for structuring the design of new studies. We are currently leveraging these taxonomies to synthesize the evidence and construct an interactive decision aid for communicators and researchers to support the development of educational and informational materials for health-related decision-making.

ACKNOWLEDGMENTS

This work was funded by the National Library of Medicine (R01 LM012964, PI: Ancker).

CONFLICTS OF INTEREST

None of the authors has any conflicts to disclose.

APPENDIX A

Table: Expanded taxonomy of data structures, with examples

				SINGLES		EFFECTS	
				Populations[1]		Pairs of populations, with and without (or pre- and post-) factor of interest[2]	
CATE-GORY	**Trend over time**	**Variable numbers, valences**		**One**	**Many**	**One**	**Many**
BASIC	**Not of interest**	**One variable, any valence**		Probability: *This patient's estimated risk of breast cancer is 4%.* Quantity. *This patient's Ha1c level is 6.5%.*	Probability: *Average risk of breast cancer in 5 countries* Quantity: *Average Ha1c levels in diabetes in 5 countries.*	Probability: *Treatment reduces cancer recurrence risk from 7% to 3%.* Quantity: *Before therapy, Ha1c level averages 8.0%. After therapy, it averages 6.5%.*[1]	Probability: *With this treatment, the relative risks of recurrence are 1.2 for older vs middle-aged patients, and 0.8 for young vs middle-aged patients.*
		Many variables, any valence		Probability: *An individual's lifetime risk of 3 cancers* Quantity: *A patient's lipid panel*	Probability: *Average risk of breast, lung, and colorectal cancer in 5 different countries* Quantity: *Average lipid values in 5 different countries*	Probability: *Smoking increases risk of lung cancer by X% and risk of oral cancers by Y%* Quantity: *This therapy would be expected to increase HDL by X mg/dL and reduce LDL by Y mg/dL*	Probability: *In younger patients, smoking raises risk by a factor of 4 for lung cancer and 5 for oral cancers. In older patients, [parallel statistics provided]* Quantity: *For men, the therapy would be expected to increase HDL by X mg/dL and reduce LDL by Y mg/dL. For women, [parallel statistics provided]*
TRADEOFFS	**Not of interest**	**Two tradeoff variables (one positive and one negative)**[2]	**For one thera-peutic option**	Probability: *With the therapy, chance of survival is 80% and chance of blood clot is 0.01%.*	Probability: *With the therapy, men's chance of survival is 80% and chance of blood clot is 0.01%. For women [parallel statistics provided].*	Probability: *With the therapy, survival chances rise from 70% to 80%, while risk of a blood clot rises from 0% to 0.01%.*	Probability: *With the therapy, men's survival chances rise from 70% to 80%, while risk of a blood clot rises from 0% to 0.01%. For women [parallel statistics provided].*
			For 2 or more thera--peutic options	Probability: *With therapy A, 80% of patients survive the cancer, but 5% will later develop a different cancer. With therapy B, survival chances are 75% but chance of the second cancer is 2%.*	Probability*: For men, the chance of survival with Therapy A is 80% and risk of blood clot is 0.01%, while the chance of survival with Therapy B is 90% and risk of blood clot is 0.05%. For women, [parallel statistics provided].*	Probability: *Therapy A increases survival chances from 70% to 80%, but also increases the chances of a blood clot from 1% to 5%. By contrast, therapy B increases survival chances from 70% to 75% but increases the chance of a blood clot from 1% to 2%.*	*Probability: For men, Therapy A increases survival from 70% to 80% and increases risk of blood clot from 0.01% to 0.02%, while Therapy B increases survival from 70% to 90% and risk of blood clot from 0.01% to 0.05%. For women [parallel statistics provided].*
TIME TRENDS	**Of interest**	**One variable, any valence**		Probability: *Cancer recurrence risk for a patient for each year over the next 10 years*	Probability: *Cancer recurrence risk over 10 years post-therapy for Black, White, and Hispanic patients*	Probability: *Difference in cancer recurrence risk over the next 10 years between those who do and do not take therapy*	Probability: *Difference in cancer recurrence risk over the next 10 years between those who do and do not take therapy, for men and for women.*

	Quantity: *The patient's systolic blood pressure every month over the past year*	Quantity: *Projected average systolic blood pressure over time for Black, White, and Hispanic patients*	Quantity: *The projected effect of medication on systolic blood pressure over 12 months of administration*	Quantity: *The projected effect of medication on systolic blood pressure over 12 months, for men and for women*
Many variables, any valence	Probability: *A patient's risk of 3 cancers over the next decade* Quantity: *An individual's systolic and diastolic blood pressure over the previous year*	Probability: *Risks of lung and colorectal cancers among patients in the US, Canada, and Mexico over a 10-year period*	Probability: *Effect of therapy on survival chances over time* Quantity: *Projected effect of therapy on individual's systolic and diastolic blood pressure over the coming year*	Probability: *Effect of therapy on survival chances for high, medium, and low risk patients over time* Quantity: *Projected effect of therapy on individual's systolic and diastolic blood pressure over the coming year, for men and women*

1. A population may be a group of any size; a population of size 1 is the patient him- or herself.
2. Pre-post comparisons when the time trend is not of primary interest are depicted in the right-hand column as "pairs of populations." When the time trend between pre-treatment and post-treatment is of primary concern, the communication is classified in the bottom rows as a time trend.
3. Quantity examples are omitted from trade-off cells for space reasons and because they are relatively rare in trade-off messaging.

References

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211.

Ancker, J. S., & Kaufman, D. (2007). Rethinking health numeracy: a multidisciplinary literature review. *Journal of the American Medical Informatics Association : JAMIA*, *14*(6), 713-721. https://doi.org/10.1197/jamia.M2464

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*(6), 608-618. https://doi.org/10.1197/jamia.M2115

Anderson, B. L., Obrecht, N. A., Chapman, G. B., Driscoll, D. A., & Schulkin, J. (2011). Physicians' communication of Down syndrome screening test results: The influence of physician numeracy. *Genetics in Medicine*, *13*(8), 744-749. https://doi.org/10.1097/GIM.0b013e31821a370f

Bailey, S. C., Pandit, A. U., Yin, S., Federman, A., Davis, T. C., Parker, R. M., & Wolf, M. S. (2009). Predictors of misunderstanding pediatric liquid medication instructions. *Fam Med*, *41*(10), 715-721.

Becker, M. H. (1974). The health belief model and personal health behavior. *Health Education Monographs*(2), 324-508.

Bonner, C., Trevena, L. J., Gaissmaier, W., Han, P. K. J., Okan, Y., Ozanne, E., Peters, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current Best Practice for Presenting Probabilities in Patient Decision Aids: Fundamental Principles. *Med Decis Making*, *41*(7), 821-833. https://doi.org/10.1177/0272989x21996328

Cuite, C., Weinstein, N., Emmons, K., & Colditz, G. (2008). A test of numeric formats for communicating risk probabilities. *Med Decis Making*, *28*. https://doi.org/10.1177/0272989x08315246

Fair, A. K., Murray, P. G., Thomas, A., & Cobain, M. R. (2008). Using hypothetical data to assess the effect of numerical format and context on the perception of coronary heart disease risk. *American Journal of Health Promotion*, *22*(4), 291-296.

Freeman, T. R., & Bass, M. J. (1992). Risk language preferred by mothers in considering a hypothetical new vaccine for their children. *CMAJ : Canadian Medical Association journal =*

journal de l'Association medicale canadienne, *147*(7), 1013-1017. https://pubmed.ncbi.nlm.nih.gov/1393896

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1336288/
Galesic, M., & Garcia-Retamero, R. (2011). Do low-numeracy people avoid shared decision making? *Health Psychology*, *30*(3), 336-341. https://doi.org/10.1037/a0022723
Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2011). Enhancing understanding and recall of quantitative information about medical risks: A cross-cultural comparison between Germany and Spain. *The Spanish journal of psychology*, *14*(1), 218-226.
Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2011). Enhancing understanding and recall of quantitative information about medical risks: a cross-cultural comparison between Germany and Spain. *Span J Psychol*, *14*(1), 218-226. https://doi.org/10.5209/rev_sjop.2011.v14.n1.19
Graham, P. H., Martin, R. M., & Browne, L. H. (2009). Communicating breast cancer treatment complication risks: When words are likely to fail. *Asia-Pacific Journal of Clinical Oncology*, *5*(3), 193-199. https://doi.org/https://doi.org/10.1111/j.1743-7563.2009.01232.x
Grimes, D. A., & Snively, G. R. (1999). Patients' understanding of medical risks: Implications for genetic counseling. *Obstetrics and Gynecology*, *93*, 910-914. http://www.sciencedirect.com/science/article/pii/S0029784498005675
Housten, A. J., Kamath, G. R., Bevers, T. B., Cantor, S. B., Dixon, N., Hite, A., Kallen, M. A., Leal, V. B., Li, L., & Volk, R. J. (2020). Does Animation Improve Comprehension of Risk Information in Patients with Low Health Literacy? A Randomized Trial. *Medical decision making : an international journal of the Society for Medical Decision Making*, *40*(1), 17-28. https://doi.org/10.1177/0272989X19890296
Keller, C., Kreuzmair, C., Leins-Hess, R., & Siegrist, M. (2014). Numeric and graphic risk information processing of high and low numerates in the intuitive and deliberative decision modes: An eye-tracker study. *Judgment and Decision Making*, *9*(5), 420-432.
Kirsch, I. S. (2001). *The International Adult Literacy Survey (IALS): Understanding What Was Measured*. ETS.
Kreuzmair, C., Siegrist, M., & Keller, C. (2016). High Numerates Count Icons and Low Numerates Process Large Areas in Pictographs: Results of an Eye-Tracking Study. *Risk Anal*, *36*(8), 1599-1614. https://doi.org/10.1111/risa.12531
Leonhardt, J. M., & Robin Keller, L. (2018). Do Pictographs Affect Probability Comprehension and Risk Perception of Multiple-Risk Communications? *Journal of Consumer Affairs*, *52*(3), 756-769. https://doi.org/https://doi.org/10.1111/joca.12185
Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37-44.
Lokker, N., Sanders, L., Perrin, E. M., Kumar, D., Finkle, J., Franco, V., Choi, L., Johnston, P. E., & Rothman, R. L. (2009). Parental Misinterpretations of Over-the-Counter Pediatric Cough and Cold Medication Labels. *Pediatrics*, *123*(6), 1464-1471. https://doi.org/10.1542/peds.2008-0854
Marteau, T. M., Senior, V., & Sasieni, P. (2001). Women's understanding of a "normal smear test result": experimental questionnaire based study. *BMJ (Clinical research ed.)*, *322*(7285), 526-528. https://doi.org/10.1136/bmj.322.7285.526
McAndrew, L. M., Musumeci-Szabo, T. J., Mora, P. A., Vileikyte, L., Burns, E., Halm, E. A., Leventhal, E. A., & Leventhal, H. (2008). Using the common sense model to design interventions for the prevention and management of chronic illness threats: from description to process [Research Support, N.I.H., Extramural

Review]. *Br J Health Psychol*, *13*(Pt 2), 195-204. https://doi.org/10.1348/135910708X295604
Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions [journal article]. *Implementation Science*, *6*(1), 1-12. https://doi.org/10.1186/1748-5908-6-42

Nagle, C., Hodges, R., Wolfe, R., & Wallace, E. M. (2009). Reporting Down syndrome screening results: women's understanding of risk. *Prenat Diagn*, *29*(3), 234-239. https://doi.org/https://doi.org/10.1002/pd.2210

Neuner-Jehle, S., Senn, O., Wegwarth, O., Rosemann, T., & Steurer, J. (2011). How do family physicians communicate about cardiovascular risk? Frequencies and determinants of different communication formats. *BMC Family Practice*, *12*(1), 15. https://doi.org/10.1186/1471-2296-12-15

Peters, E., Hart, P. S., & Fraenkel, L. (2011). Informing patients: The influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making*, *31*, 432-436.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407-413.

Pighin, S., Savadori, L., Barilli, E., Galbiati, S., Smid, M., Ferrari, M., & Cremonesi, L. (2015). Communicating Down syndrome risk according to maternal age: "1-in-X" effect on perceived risk. *Prenat Diagn*, *35*(8), 777-782. https://doi.org/10.1002/pd.4606

Reyna, V. F. (2008). A theory of medical decision-making and health: Fuzzy trace theory. *Medical Decision Making*, *28*, 850-865.

Reyna, V. F., Nelson, W. L., Han, P., & Dieckmann, N. (2009). How numeracy influences risk comprehension and medical decision-making. *Psychological Bulletin*, *135*(6), 943-973.

Rhodes, R. E., & de Bruijn, G. J. (2013). How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. *Br J Health Psychol*, *18*(2), 296-309. https://doi.org/10.1111/bjhp.12032

Schwartz, L., Woloshin, S., Black, W., & Welch, H. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966-972.

Siegrist, M., Orlow, P., & Keller, C. (2008). The effect of graphical and numerical presentation of hypothetical prenatal diagnosis results on risk perception. *Med Decis Making*, *28*(4), 567-574. https://doi.org/10.1177/0272989x08315237

Smerecnik, C. M. R., Mesters, I., Kessels, L. T. E., Ruiter, R. A. C., De Vries, N. K., & De Vries, H. (2010). Understanding the Positive Effects of Graphical Risk Information on Comprehension: Measuring Attention Directed to Written, Tabular, and Graphical Risk Information. *Risk Analysis*, *30*(9), 1387-1398. https://doi.org/https://doi.org/10.1111/j.1539-6924.2010.01435.x

Trevena, L. J., Bonner, C., Okan, Y., Peters, E., Gaissmaier, W., Han, P. K. J., Ozanne, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current Challenges When Using Numbers in Patient Decision Aids: Advanced Concepts. *Med Decis Making*, *41*(7), 834-847. https://doi.org/10.1177/0272989x21996342

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

van der Bles, A. M. v. d. L., Sander; Freeman, Alexandra LJ; Mitchell, James; Galvao, Ana B; Zaval, Lisa; Spiegelhalter, David J. (2019). Communicating uncertainty about facts, numbers, and science. *Royal Society Open Science*, *6*(181870). https://doi.org/10.1098/rsos.181870

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). Engineering Psychology and Human Performance.

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, *59*, 329-349.

Yin, H., Mendelsohn, A. L., Wolf, M. S., & et al. (2010). Parents' medication administration errors: Role of dosing instruments and health literacy. *Archives of Pediatrics & Adolescent Medicine*, *164*(2), 181-186. https://doi.org/10.1001/archpediatrics.2009.269

Zikmund-Fisher, B., Smith, D., Ubel, P., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale (SNS): Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making*, *27*. https://doi.org/10.1177/0272989x07303824

Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2007). Mortality versus survival graphs: improving temporal consistency in perceptions of treatment effectiveness. *Patient Educ Couns*, *66*(1), 100-107. https://doi.org/10.1016/j.pec.2006.10.013

Zikmund-Fisher, B. J., Scherer, A. M., Witteman, H. O., Solomon, J. B., Exe, N. L., Tarini, B. A., & Fagerlin, A. (2017). Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *Journal of the American Medical Informatics Association*, *24*(3), 520-528. https://doi.org/10.1093/jamia/ocw169