Risk and Uncertainty Communication

David Spiegelhalter

Centre for Mathematical Sciences, University of Cambridge, CB3 0WB Cambridge, United Kingdom; email: d.spiegelhalter@statslab.cam.ac.uk

Annu. Rev. Stat. Appl. 2017. 4:31–60

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

This article's doi:
10.1146/annurev-statistics-010814-020148

Keywords

psychology, gambling, graphics, public engagement, evidence, climate change, reliability

Abstract

This review briefly examines the vast range of techniques used to communicate risk assessments arising from statistical analysis. After discussing essential psychological and sociological issues, I focus on individual health risks and relevant research on communicating numbers, verbal expressions, graphics, and conveying deeper uncertainty. I then consider practice in a selection of diverse case studies, including gambling, the benefits and risks of pharmaceuticals, weather forecasting, natural hazards, climate change, environmental exposures, security and intelligence, industrial reliability, and catastrophic national and global risks. There are some tentative final conclusions, but the primary message is to acknowledge expert guidance, be clear about objectives, and work closely with intended audiences.

INTRODUCTION

Background

The phrase "risk communication" appears to have been used for the first time in 1984 (Leiss 1996), with the concept arising primarily from environmental concerns in the United States in the 1970s and 1980s around the Three Mile Island nuclear accident, nuclear waste disposal, siting of chemical plants, and so on. In the intervening three decades, there has been intense, and often contested, work on this topic. Environmental issues remain important, and there has been increased attention both to natural hazards such as floods, hurricanes, and climate change, and emerging technologies such as genetically modified organisms and nanotechnology. However, although such societal risks can dominate headlines, recent developments in risk communication have primarily focused on specific individuals, particularly risks to health from pharmaceuticals, genetics, and lifestyle. The example in the sidebar Is Bacon as Dangerous as Cigarettes? shows there is still much work to do.

There is a vast literature on potential risks, from mundane choices about our diet to global catastrophic threats that could spell the end of humanity. So in this article I selectively examine the communication of statistical analysis on risks to nonspecialist audiences, and carefully avoid the vital but contested topics of risk management, governance, and regulation, as well as considerations of what constitutes an acceptable risk. The examples are mainly drawn from the United States and United Kingdom.

But we first need to consider what we mean by the ill-defined terms in the title: risk, uncertainty, and communication.

What Is Risk?

In everyday English usage, "risk" generally refers to undesirable things that might happen. In contrast, the official ISO 31000 (ISO 2009) definition of risk is the more balanced, if slightly vacuous, "effect of uncertainty on objectives."

A vital distinction is between hazard, which is the potential for harm (e.g., being 6 miles up in the air in a machine is hazardous), and risk, which relates to the actual potential for harm after

IS BACON AS DANGEROUS AS CIGARETTES?

In November 2015 the World Health Organization's (WHO's) International Agency for Research in Cancer (IARC) announced the finding of an expert group that processed meat was a Group 1 carcinogen, putting it in the same category as cigarettes and asbestos. This led to headlines such as "Bacon, ham and sausages have the same cancer risk as cigarettes warn experts" (Kirby 2015). IARC desperately tried to sort out the subsequent confusion by pointing out that the Group 1 classification was about the confidence of an increased risk of cancer existing, and said nothing about the magnitude of the risk.

IARC estimated that 50 g of processed meat a day was associated with an increased risk of bowel cancer of 18%. Statistical commentators then reframed this relative risk into a change in absolute risk (Goldin 2015): In the normal run of things, around 6 in every 100 people would be expected to get bowel cancer in their lifetime, but if 100 similar people ate a bacon sandwich every single day of their lives, then according to the IARC report we would expect that 18% more would get bowel cancer—which is a rise from 6 cases to 7 cases. That is one extra case of bowel cancer in all those 100 lifetime bacon-eaters, which puts things into perspective.

mitigating actions have been taken (planes are now amazingly safe). Risk is often considered as a function of both the probability of an adverse outcome and the magnitude of the consequences, and in the domain of safety, natural hazards, and disasters, risk = hazard × vulnerability × exposure (Geosci. Aust. 2014). Scheer et al. (2014) show that precise definitions vary widely between major organizations: For example, the US Environmental Protection Agency (EPA) defines hazard as "a potential source of harm," whereas risk is narrowly defined as "the probability of adverse effects resulting from exposure to an environmental agent or mixture of agents" (EPA 2011).

From a theoretical perspective, rational decision-making in the face of risk comprises four basic stages (Savage 1951):

1. structuring the list of actions, and the possible consequences of actions,
2. giving a value to those possible futures,
3. assigning a probability for each possible consequence, given each action, and
4. establishing the rational decision as that which maximizes the expected benefit.

This theory holds in situations of perfect contextual knowledge, such as gambling on roulette when we know the probabilities and possible gains or losses. These perfect-chance problems are the only type in which we might talk about a truly known risk, and even then we need to make assumptions about the fairness of the wheel and the integrity of the casino. Savage (1972) introduced the idea of an idealized small world with which a statistical model deals, but the real, big world is a far messier place, in which we do not have access to all the information necessary to carry out this normative rational behavior. So risk assessment, and therefore risk communication, has to deal with this additional uncertainty.

What Is Uncertainty?

The answer to this question depends on whom you ask. Many scientists would use the term "uncertainty" for everything that is not certain, including a single coin flip, and only distinguish the extent to which the uncertainty is quantifiable. In contrast, those with an economics and social science background often adopt the distinction (Knight 1921) between risk, in which there is agreed quantification due to extensive data and good understanding of a controlled environment, and uncertainty, for when this is not feasible. The term "ambiguity" is used in behavioral economics to refer to uncertainty about probabilities (Ellsberg 1961) but, rather ironically, the term is ambiguous, as others use it to refer to situations in which outcomes and values are contested or unknown (Stirling 2007).

To try and resolve this conundrum, there have been a multitude of different attempts at taxonomies for structuring uncertainty that runs deeper than simple chance variation, many driven by integrated assessment models in climate modeling (see Galesic et al. 2016 for an encompassing review). Perhaps the simplest three-level categorizations have been provided by Donald Rumsfeld's (much derided at the time) "known knowns, known unknowns, and unknown unknowns" (Rumsfeld 2002); known, unknown, and unknowable in economics (Diebold et al. 2010); or simply risk, uncertainty, and ignorance.

Sticking to this three-level categorization, but focusing on statistical modeling of risks, it is useful to think in terms of the following:

- *Aleatory uncertainty*: inevitable unpredictability of the future due to unforeseeable factors, fully expressed by classical probabilities

- *Epistemic uncertainty*: uncertainty about the structure and parameters of statistical models, expressed, for example, through Bayesian probability distributions, default parameter values, safety factors, and sensitivity analyses to assumptions (Morgan et al. 2009). The crucial aspect is that this is still modeled and quantified uncertainty.
- *Ontological uncertainty*: uncertainty about the entire modeling process as a description of reality. By definition, this is not part of the modeled uncertainty, and can only be expressed as a qualitative and subjective assessment of the coverage of the model, conveying with humility the limitations of our knowledge.

When it comes to communicating the consequences of aleatory and epistemic uncertainty, the standard tools are to provide point estimates of risks, accompanied by distributions, ranges, or a list of alternatives driven by sensitivity analyses. The final, unmodeled uncertainty is more challenging, and some application areas have adopted summary qualitative scales to communicate the confidence in their modeling.

Such doubts about the whole modeling process arise from scientific uncertainty due to limitations in the available evidence, and this presents a complex communication challenge (Fischhoff & Davis 2014). But it is clear that whatever the approach taken to facing up to deeper uncertainties, the result is to cast doubt on any optimal analysis driven by classical decision theory, and this doubt leads to decision-making that is robust to known inadequacies in the analysis, and resilient to unknowns (Lempert & Collins 2007). A natural consequence is for risk managers to take a precautionary approach, although this arises naturally as a holding strategy in the face of inadequate understanding, rather than as an overriding principle (Stirling 2007).

What Is Good Communication?

An excellent resource is the guide from the Food and Drug Administration (FDA) Risk Communication Advisory Committee (Fischhoff et al. 2011), which begins with the quote "A risk communication is successful to the extent that it contributes to the outcomes its sponsor desires" (NRC 1989, p. 3). This may seem rather evasive, but serves to emphasize that we cannot assess the quality of risk communication unless the objectives are clear. Outcome measures used in studies of risk communication can include audiences' comprehension, their liking the material, feelings aroused, attitudes regarding threats, intentions for the future, and (rarely) actual changes in behavior. More subtly, we may seek to generate "immunity to misleading anecdote" to reduce the impact of idiosyncratic stories garnered from acquaintances and social media (Fagerlin et al. 2005). Crucially, we need to be clear about whether we are seeking to persuade, or fulfilling a duty to inform. I assume the latter throughout this review.

It is implicit in the preceding definition of successful risk communication that we need to properly understand the perceptions and feelings of the audience, but this has not always been recognized. Three main phases in the history of risk communication were identified by Leiss (1996): Phase I was identified with expertise, in that experts produced numerical risk assessments based on the best available knowledge, and felt that simply communicating these to the public in a clear way would lead people to agree with the rational decisions being proposed. The idea that public skepticism is due to their lack of understanding, and that more education would resolve conflicts, was christened the "deficit model" by UK social scientists in the 1980s (Wynne 1991), and in the United States, the phrase "public irrationality thesis" has been similarly used in a derogatory sense. Leiss identified 1984 as the end of this phase (although in my personal experience many seem not to have progressed past it).

The idea that the public were irrational about risk due to their misperceptions was strengthened by work in the 1970s and 1980s around technology anxiety, summarized in a classic review article

(Slovic 1987). Working within a psychometric paradigm (Slovic 2000), psychologists found that the following factors were associated with increased perception of the magnitude of, and anxiety about, the risks of a particular hazard:

- Uncontrollable
- Having catastrophic potential
- Having fatal or dread consequences
- Bearing an inequitable distribution of risks and benefits
- Not understood
- Novel
- Delayed in their manifestation of harm

These have been termed fear factors (Ropeik 2010), and it is instructive to see how many boxes are ticked by a hazard such as, for example, escaped radiation after a nuclear accident. Furthermore, when a crisis arises, Peter Sandman has argued that the public's response to risks is driven not only by the nature of the hazard, but also their outrage that such events should happen, accompanied by a strong urge to blame someone (Sandman et al. 1993, Lachlan & Spence 2010). These insights led to Leiss's Phase II, termed trust, based on the belief that people will only accept appropriate risks if they can be persuaded to trust the source of information.

But humans are not so easily malleable, and more recent research has revealed insights that require us to go beyond the idea that public responses are due to errors and biases that might be remedied by information from an acknowledged authoritative source. An important idea is the affect heuristic, which is essentially that people tend to have an overall positive or negative feeling about a potential hazard (Slovic & Peters 2006): Negative affect, say toward anything nuclear, means they will tend to both play up the harms and downplay any benefits, whatever expert sources might say.

Another powerful innovation is derived from the cultural theory of risk, in which attitudes of individuals to risk are seen as arising from their beliefs about the principles behind an ideal society (Douglas & Wildavsky 1983). A psychometric approach to this anthropological idea has been brought together in the cultural cognition framework, or worldviews, approach: For example, Kahan (2015) found a strong correlation between political opinion and beliefs in the risk from climate, gun possession, and fracking, but no relation when asked about, say, radio waves from cellphone towers or nanotechnology. Although the theoretical basis for the cultural cognition thesis has been criticized (van der Linden 2016), these studies provide further explanation for why increased knowledge may not, for some culturally specific issues, be correlated to feelings about risks.

This leads us to Leiss's Phase III, which is based on the idea of authorities becoming trusted by demonstrating trustworthiness. The aim is to develop social trust (Cvetkovich & Lofstedt 2013) by not just trying to persuade, but engaging in genuine mutual understanding and strategic listening (Pidgeon & Fischhoff 2011). Such understanding requires a more sophisticated segmentation of the public (or publics), and one suggestion is to create a range of mental models, which display as directed graphs different conceptualizations of the issue in question (Morgan et al. 2001). Such an approach is particularly relevant in highly contested areas, such as nuclear waste disposal in Sweden (Skarlatidou et al. 2012).

Even in areas that are not so controversial, messages need to take into account the wide variability in individual characteristics such as numeracy and ability to understand graphs (Peters 2008). Both in the planning and throughout the process of development of risk communication materials, close cooperation and even coproduction with potential user groups is the key.

The sidebar Developmental Stages in Risk Communication provides a terse and somewhat ironic summary.

DEVELOPMENTAL STAGES IN RISK COMMUNICATION (FISCHHOFF 1995)

1. All we have to do is get the numbers right
2. All we have to do is tell them the numbers
3. All we have to do is explain what we mean by the numbers
4. All we have to do is show them they've accepted similar risks in the past
5. All we have to do is show them it's a good deal for them
6. All we have to do is treat them nice
7. All we have to do is make them partners
8. All of the above

It has been over twenty years since Fischhoff laid out these stages, but we still appear to be struggling through the list.

Structure of This Review

It is futile to try to cover practice across the wide range of domains in which risks are communicated, so I begin by exploring the major issues with a specific focus: communicating individual health risks. This is the area that has been most widely researched, and illustrates current thinking concerning communicating probabilities with numbers, words, and graphics, and also dealing with deeper uncertainties. I then cover a range of case studies in areas ranging from climate change to security and intelligence, describing how they deal with the common challenges, and whether the focus is individual people, specific events, societal risks, and even global risks. I avoid discussing risk communication at times of acute crises, when numerical assessments would generally be inappropriate. I also avoid financial risk communication, which requires its own review.

There are a huge number of sources for advice on risk communication, for example, the US EPA (**https://www.epa.gov/risk/risk-communication#self**) and the UK Cabinet Office (UK Cabinet Off. 2011), although they generally give limited attention to communicating numbers. Even in their authoritative review, Fischhoff et al. (2011) say they can only give a best guess at best practice—any communication material always needs careful evaluation in the particular circumstances to see if it is likely to fulfill its objectives. So although I hope this review might improve sharing of good practice between different application areas, my final summary will be very tentative.

COMMUNICATING PROBABILITIES OF EVENTS

Recent research in quantitative risk communication has largely concerned risks to individuals from medical interventions, their lifestyle or their genes: systematic reviews include Visschers et al. (2009) and Zipkin et al. (2014), who listed 91 studies including 74 randomized trials, using outcomes ranging from comprehension of information, affective states, or behavioral outcomes such as decisions or intention to change behavior. Good checklists have been provided (Fagerlin et al. 2011), and Trevena et al. (2013) provided a very valuable expert consensus on good practice for including quantitative information in decision aids for patients.

It helps to be able to refer a personal example. **Figure 1** shows the way in which 23andMe (**https://www.23andme.com/en-gb/**) communicated a risk to me based on a submitted saliva sample.

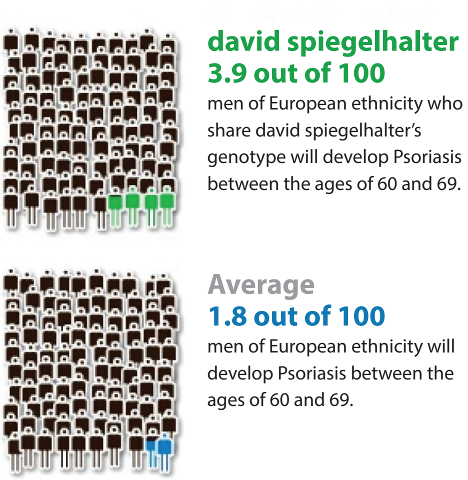

Figure 1

The form in which 23andMe communicated my risk of developing psoriasis between 60 and 69 (I am 63), derived from a limited set of genetic markers. Image generated by **https://www.23andme.com/en-gb/**.

Positive and Negative Framing

The behavioral economic work of Kahneman and Tversky (Kahneman et al. 2006) has shown that numbers do not speak for themselves—how they are expressed can have an impact on our feelings, in particular whether they are framed in a positive or negative way. For example, in the United States there is roughly a 2% *mortality* rate from isolated coronary artery bypass surgery (Society of Thoracic Surgeons 2015), whereas in the United Kingdom you can find the *survival* rate of your named cardiac surgeon, with an average of 98% (NHS Choices, **https://www.nhs.uk/Service-Search/consultants/performanceindicators/1018**). Multiple experiments have shown that a positive frame (e.g., survival) makes potential harms look less and can increase acceptability of therapies: For example, Peters et al. (2011) found in a randomized experiment that patients told that "90% of patients do not get a bad blistering rash" perceived a medication as less risky than those told that "10% of patients get a bad blistering rash" (1.43 versus 1.82 on a scale of 1 to 5).

If we want to avoid deliberate positive or negative framing, then both might be provided (Gigerenzer 2014), helped by a clear part-to-whole comparison, as in the portrayal in **Figure 1** of all 100 people who will or will not develop psoriasis.

Ways of Expressing Chances

There has been a long dispute over whether it is better to express a proportion as a percentage, say, 20%, or as a frequency, say, 20 out of 100 people like you, as in **Figure 1**. No clear preference has emerged, although in one randomized experiment with 298 participants, Peters et al. (2011) did find an interaction with numeracy: People with lower numeracy perceived a percentage as less risky than the corresponding frequency, but high-numeracy subjects did not.

Two advantages of a frequency format are its suitability to representation as icon arrays (pictograms) as in **Figure 1**, and its suitability for chances <1% when percentages require decimals. However, it is difficult to give a frequency interpretation to truly unique events, such as the future prospects for the planet (although one can perhaps envisage 100 possible futures for the earth).

GETTING THE REFERENCE CLASS RIGHT

In 2016 a UK survey led to headlines such as "Why are only 1 in 100 men taking up shared parental leave?" (Kemp 2016). Close examination of the story revealed that the survey included all men in the denominator, not just those eligible for paternity leave, so it was hardly surprising the proportion was low. In this case the reference class was clearly stated, but was completely inappropriate.

More important than the choice of format is being absolutely clear as to what the probability actually means (Morgan et al. 2009), which requires careful specification of the reference class (Gigerenzer & Galesic 2012). For example, when communicating the percent chance of rain, does this mean the fraction of the time it rains, the fraction of the area on which it rains, or the fraction of situations like this in which it rains? The last is the correct interpretation. This can favor a frequency format, because the denominator is then explicitly described—see the sidebar Getting the Reference Class Right.

Trevena et al. (2013) also emphasize the need to make the time interval clear, and not to say "10% chance of *X*," but either "10 out of 100 people each year" or "Every year, 10% of people." And whatever the choice between percentages and frequencies, Tait et al. (2010) found that understanding is better when presented in a table rather than just included in text.

Chances as Fractions

If using a frequency format, a decision has to be made whether, for example, to say "5 out of 100," "50 out of 1,000," and so on. The influence of this choice is known as ratio bias, and it is generally argued that a larger numerator suggests a larger risk: A classic study showed that participants rated cancer as riskier when it was described as killing 1,286 out of 10,000 people than as killing 24.14 out of 100 people (Yamagishi 1997). The extreme case of ratio bias occurs when the denominator is entirely ignored, and this is termed denominator neglect—when the media demand precautionary measures following a tragedy involving a child, the vast number of safe journeys are not considered. A frequency format with a clear reference class can counter this (Garcia-Retamero et al. 2010).

A common format for risks is "1 in *X*," which the preceding argument suggests should tend to make risks look smaller than, say, "10 out of 10*X*." This format has been strongly criticized (Zikmund-Fisher 2014), although a recent review argued that this effect is smaller than claimed (Sirota et al. 2014), and a recent study found that "1 in *X*" can even be seen as suggesting higher risks than when expressed with a higher numerator (Pighin et al. 2015), particularly for participants with low educational level.

Comparative Risks

Table 1 illustrates a range of comparative measures applied to the example described in the sidebar Is Bacon as Dangerous as Cigarettes?

There is strong agreement that comparing risks that are all expressed in "1 in *X*" format is inappropriate (Zikmund-Fisher 2014): In a classic survey, Galesic & Garcia-Retamero (2010, p. 464) asked, "Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1,000, or 1 in 10?" Only 72% of 1,000 respondents in the United States and 75% of 1,000 respondents in Germany answered correctly.

Table 1 Examples of methods for communicating the lifetime risk of bowel cancer with and without a daily bacon sandwich

Method	Non–bacon eaters	Daily bacon eaters
Event rate	6%	7%
Expected frequency	6 out of 100	7 out of 100
Absolute risk difference		1% or 1 out of 100
"1 in *X*"	1 in 16	1 in 14
Relative risk		18%
NNT		100

Abbreviation: NNT, the number needed to treat, which in this case is number of people needing to eat a bacon sandwich every day of their lives, in order to expect one extra case of bowel cancer (and so perhaps better defined as the "number needed to eat"). Expected frequency and absolute risk difference are recommended.

The number needed to treat—the inverse of the difference in absolute risks—is also not recommended: Zipkin et al. (2014) concluded that it is neither well-understood nor popular with patients, which is unsurprising because the concept is relevant to cost-effectiveness analysis of policies rather than individual decisions.

Systematic reviews have concluded that using relative measures increases the perceived size of the effect and decreases understanding (Akl et al. 2011), although the various studies used a wide variety of information (Covey 2007), and Sorensen et al. (2008) concluded that if the baseline is clearly given, the difference between relative and absolute risk largely disappears. We will also see later that, for low-probability, high-impact events, a relative risk can be valuable.

Trevena et al. (2013) concluded that risks should either be compared with simple percentages or use a frequency format and keep the denominator fixed. A frequency format also allows the incremental risk to be clearly communicated, for example, as "one extra case of bowel cancer expected in every 100 people eating processed meat daily." The BBC used this in their news bulletins following the WHO processed meat story. But a trial suggested this format worked best when accompanied by a graphical display (Zikmund-Fisher et al. 2008).

Graphics

There have been many systematic and nonsystematic reviews of the use of graphics in risk communication, primarily in the health fields (Ancker et al. 2006, Visschers et al. 2009, Spiegelhalter et al. 2011, Hildon et al. 2012, Stephens et al. 2012, Trevena et al. 2013, Zipkin et al. 2014, Hallgreen et al. 2016). Rather limited conclusions have been drawn, with authors emphasizing the overwhelming need for proper evaluation and testing, and the formats' crucial dependence on the audience and the intention: "graphical features that improve accuracy of quantitative reasoning appear to be different from features that induce behavior change, and features that viewers like may not support either of the two goals" (Ancker et al. 2006, p. 609). Graphics also may not help those with low graphical literacy (Gaissmaier et al. 2012).

There is agreement that graphs can be effective in conveying the gist of the message, whereas numerical formats are appropriate for more detailed, verbatim information (Feldman-Stewart et al. 2007). Line graphs are good for trends, and bar charts are good for comparisons between groups, as the viewer can get the gist very quickly. Showing only adverse events can lead to risk-aversion because of their foreground salience (Stone et al. 2003), and so to avoid these framing effects one would ideally show the part-to-whole comparison of both the positive and negative

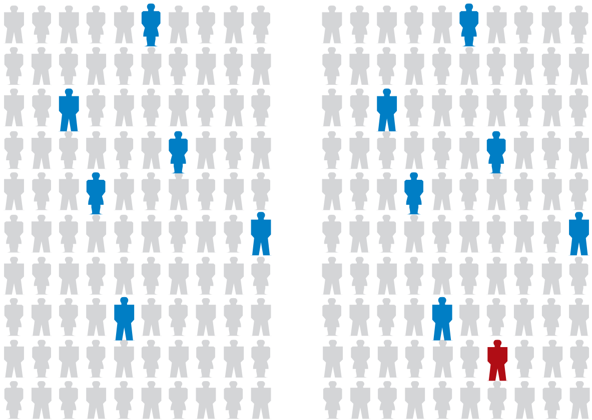

Figure 2

Bacon sandwich example: paired arrays, scattered icons, incremental risk format. Of 100 people who do not eat bacon, 6 (*blue icons*) develop bowel cancer in the normal run of events. Of 100 people who eat bacon every day of their lives, there is 1 additional (*red*) case.

outcomes. Although pie charts do this, they are very difficult to compare and have been broadly criticized. Risk scales over many orders of magnitude are discussed below.

Icon arrays (pictographs) have been more particularly investigated, and shown to be effective in improving the understanding of people with low numeracy (Galesic et al. 2009, Galesic & Garcia-Retamero 2010). **Figures 1** and **2** both compare risks, but in the former the icons have been blocked, while for the processed meat example the icons have been scattered and displayed in an incremental format. Rather intuitively, scattering increases the impression of randomness and unpredictability, but makes it very difficult to count, and so is considered poor for comparisons (Ancker et al. 2011). An eye-tracking study showed that people process icon arrays in different ways—those with high numeracy tend to count icons, whereas those with low numeracy evaluate area, so icons need to be blocked to do this (Kreuzmair et al. 2016).

Few (2016) argued strongly that icons are inappropriate if they cannot quickly give the gist, and that audiences should neither have to count icons nor compare areas—this is a particular problem if a number of different outcomes are being represented. He suggested using either bar charts or icons blocked into bars allowing visual comparisons by height alone, rather than displaying in a regular grid of, say, 10×10 (although **Figure 2** might be reasonable as it is a simple case with an increment of 1 that does not require counting). When it comes to choosing the number of icons, Zipkin et al. (2014) claimed that the denominator should be 1,000, but the evidence for this is not strong, and **Figure 1** shows an effective use of partially colored icons.

It is tempting to make graphics overcomplex and, under the label less-can-be-more, Zikmund-Fisher et al. (2010) showed how simplifying a multi-outcome array to just two outcomes improved understanding: As an extreme example, the popular Your Disease Risk website (Washington University, **http://www.yourdiseaserisk.wustl.edu/**) only uses colored relative risk levels to communicate the gist and avoids numbers altogether. In the same vein, modern technology allows superficially attractive interactive animations, but Zikmund-Fisher et al. (2011, p. 1) claimed these can be "cool but counter-productive" and any unnecessary complexity, whether through animation or scattered icons, can degrade performance (Zikmund-Fisher et al. 2012).

This again emphasizes the difficulty of producing rules, that one size does not fit all, and developers must use good design sense and test all material with a variety of audiences.

Table 2 Form of words for drug adverse events mandated by the European Medicines Agency (EMA 2016) for information inserts in drug packaging

Verbal Description	Frequency group
Very common	≥1/10
Common	≥1/100 to <1/10
Uncommon	≥1/1,000 to <1/100
Rare	≥1/10,000 to <1/1,000
Very rare	<1/10,000
Frequency not known (cannot be estimated from the available data)	

Using Words

Using words alone to communicate numerical risks is generally discouraged. Although some people may prefer to receive information as words rather than numbers (Trevena et al. 2006), numbers are trusted more and preferred by many (Visschers et al. 2009). Furthermore, using a qualitative descriptor such as "high risk" alone, without a numerical interpretation, can lead to lower accuracy and satisfaction, and higher risk aversion (Zipkin et al. 2014).

Words may also be used as an adjunct to numerical risks as evaluative labels to provide context to a number, although Trevena et al. (2013) recommended caution, as they can have a substantial effect through arousal of affect. Within moderation, this can be an advantage: Peters et al. (2009) showed that subjects who were randomized to see labels characterizing the percentage of pneumonia patients who survive while being treated in one of the categories of poor, fair, good, or excellent showed greater use of the numbers, and suggested less influence of their current mood, at least among the less numerate.

The interpretation of descriptive labels may not, however, be as intended. **Table 2** shows the scale dictated by the European Medicines Agency (EMA) for use on patient information leaflets in the United Kingdom. The definition of "common" as "1% to 10%" may match the usage of pharmacologists, but in a randomized trial of 120 patients, the mean estimate of the likelihood of a common side effect was 34% (Knapp 2004). The EMA now recommends using both words and frequency bands, for example, "Common: may affect up to 1 in 10 people," but including words still tends to increase the perceived risk (Knapp et al. 2016).

Visschers et al. (2009) concluded that people claim to prefer to receive information in numerical form, trusting it more, but pass it on using evaluative words. They recommend using both words and numbers, which we shall later see in other examples.

Communicating Chronic Risks

A wide range of different metrics and metaphors have been adopted when trying to communicate chronic risks of adverse events that may occur well into the future (Trevena et al. 2013). These include:

1. *Hazard ratios*. These measure the relative increase in the risk of an event in a fixed period of time, say a year, and arise naturally from epidemiological studies. But they suffer from all the disadvantages of relative risks, and alternative formats using absolute risks are preferable, as in the processed meat example.
2. *Absolute risks of events at a fixed time in the future*. For example, QRISK2 might, on the basis of the risk factors reported by an individual, conclude, "Your risk of having a heart attack or stroke within the next 10 years is 12%. In other words, in a crowd of 100 people with the same risk factors as you, 12 are likely to have a heart attack or stroke within the next

10 years." accompanied by a scattered icon array (**http://www.qrisk.org/**). Risk calculators in the United States are based on the Framingham data and could report, for example, "Risk score: 12%. Means 12 of 100 people with this level of risk will have a heart attack in the next 10 years" (National Heart, Lung and Blood Institute, **http://cvdrisk.nhlbi.nih.gov/**). Both use a mixture of percentage and frequency formats.

3. *Cumulative lifetime risks*. These are calculated under different assumptions, allowing eye-catching statements such as "1 in 2 people born after 1960 in the United Kingdom will be diagnosed with some form of cancer during their lifetime" (Cancer Research UK 2015).
4. *Survival curves*. These may appear fairly complex, but can be presented at selected times points, and a study showed that a short tutorial enabled people to accurately answer questions from full survival curves (Rakow et al. 2012).
5. *Changes in life expectancy*. "Years off your life" is a popular metric for, say, communicating the effects of smoking. However, Harmsen et al. (2014) found that of patients randomized to be told, for example, that if 100 people like them took statins for 10 years, then one fewer would die, then 25% accepted statins, whereas of those randomized to be told that taking statins would on average add 4 months to their life, only 5% accepted statins. So the absolute risk format was far more influential, possibly since the gain in life appeared unimpressive and also delayed until old age.
6. *Changes in effective age*. An increasingly popular metric might be termed effective age, the age of an average healthy person with the same level of risk as you. There are many commercial websites willing to calculate your so-called real age, and Heart Age calculators are widely used (NHS Choices, **https://www.nhs.uk/tools/pages/heartage.aspx**) and in a small trial produced a greater improvement in risk factors than a risk score alone (Lopez-Gonzalez et al. 2015).
7. *Time lost per exposure*. Phrases such as "each cigarette takes 15 minutes off your life" are vivid but must be considered as metaphors (Spiegelhalter 2012).
8. *Comparators*. Metrics can be used to provide comparative measures. For example, assuming a hazard ratio of 1.06 in all-cause mortality per 10 $\mu g/m^3$ of PM2.5 (COMEAP 2009), living in Delhi with an annual average concentration of 122 $\mu g/m^3$ (WHO 2016) gives a rough hazard ratio of 2.0, or around that of 20 cigarettes a day. So the effects of air pollution can be transformed to "cigarette equivalents."

This wide range of alternative metrics presents a wide choice, but there is a pressing need for further studies to explore the appropriate contexts for their use.

Communicating Small Risks and Using Comparators

There are particular problems in communicating and comparing small risks, because people have trouble discriminating between, say, 1 in 10,000 and 1 in 100,000. A common solution is to use a logarithmic risk scale from certainty down to, for example, 1 in 10,000,000—if drawn vertically, these are known as risk ladders. A popular example is the Paling scale (Paling 2003), which uses minimal, low, very low, etc. as descriptors, although note previous warnings about evaluative labels.

The design of the scale is vital, as perceived risk is associated with position rather than the underlying magnitude (Ancker et al. 2006, Sandman et al. 1994). There is also the problem of encouraging perception on a relative scale, so that a change in risk from 1 in 10 to 1 in 100 might be seen as equivalent to a change from 1 in 1,000 to 1 in 10,000.

Inclusion of comparators, such as smoking on a scale of radon exposure, has been shown to help understanding in people with low numeracy skills (Keller et al. 2009). However, Visschers et al. (2009) warned that comparisons need to be carefully chosen so as not to arouse inappropriate affect, such as comparisons with acknowledged risky activities such as smoking or riding a motorbike. A common choice is the annual chance of being struck by lightning, which happens to around 1 in 1,000,000 people in the United States per year. But Fischhoff (2006) pointed out that this event has special qualities that may influence perceptions—it is newsworthy and hence might seem more likely, it varies widely between urban and rural dwellers (or golfers), and it is an acute and violent threat.

This suggests a range of possible comparators measured on a common scale, such as the micromort, which is a 1 in 1,000,000 chance of death: This allows comparisons of everything from serving in a war to heart surgery, hang-gliding, and heroin use (Spiegelhalter 2014).

COMMUNICATING UNCERTAINTY ABOUT NUMERICAL RISKS

In the next two sections we consider epistemic uncertainty, also known as second-order uncertainty, about the numerical risks in a modeling framework, and ontological uncertainty about the whole modeling process due to limited scientific understanding.

When it comes to expressing epistemic uncertainty about the numerical risks, a hierarchy of levels of precision can be assigned, for example:

1. *Numbers given to appropriate levels of precision*. This is usually reflected by the number of significant digits (for example, saying 10% rather than 10.2%) or rounding (saying 10% rather than 12%), and this applies when using frequencies (1 out of 5 conveys less precision than 22 out of 100).
2. *A distribution or range*, for example, a confidence interval. Unfortunately, when confronted with a range, many people perceive the underlying distribution as uniform, although more numerate individuals are more likely to perceive the distribution as roughly normal (Dieckmann et al. 2015). There is some evidence that explicitly communicating uncertainty about a risk can increase anxiety owing to ambiguity aversion, although this might be moderated by optimism (Han et al. 2011, Trevena et al. 2013).
3. *A measure of statistical significance* compared with a null hypothesis. This might be a standard p-value, but can also be expressed in terms of the number of standard errors from the null. For example, the funnel plot shown in **Figure 3** is a popular means of displaying estimates of past risk faced by patients, as well as boundaries indicating statistical significance at the 5% and 0.2% level, which for a Normal approximation corresponds to two and three standard errors from the average performance. Patients can readily interpret such plots with some helpful explanation, although the central line is perhaps best not included to avoid excessive attention to whether a point is above or below the mean (Rakow et al. 2015). The translation of small p-values into the corresponding number of standard errors, as in the confirmation of the Higgs Boson with a five-sigma result (van Dyk 2014), is an attractive shorthand.
4. *Verbal qualifiers to numbers*. New UK patient-information leaflets clearly report the numbers of women who would be expected to benefit and be harmed out of each 200 attending breast screening (NHS Breast Screening Programme, **https://www.gov.uk/government/publications/breast-screening-helping-women-decide**). These numbers relied to a considerable extent on expert judgment, and their pedigree was reported using the phrase "The numbers ... are the best estimates from a group of experts who have reviewed the evidence" (p. 10). A verbal descriptor can also be given for a range, although this raises issues discussed earlier.
5. *A refusal to give a number* unless the evidence is good enough (see below in the section Climate Change).

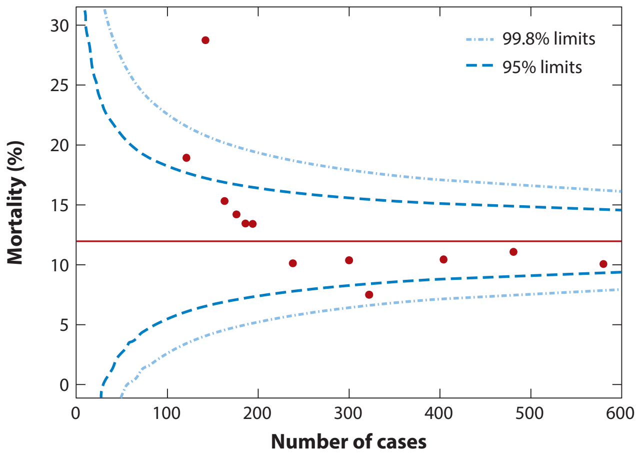

Figure 3

A funnel plot of surgical mortality in heart surgery in children under the age of one year in 12 UK centers between 1991 and 1995: The funnels correspond to two and three standard errors. This simple plot manages to display the estimated risks and boundaries for statistical significance, and it reveals a clear volume effect (Spiegelhalter 2002).

COMMUNICATING CONFIDENCE IN THE ANALYTIC PROCESS

When it comes to communicating limited scientific understanding, there have been numerous suggestions for expressing overall confidence in an analysis based on the quality and strength of the evidence. Approaches have been reviewed by Spiegelhalter & Riesch (2011), and include the following:

- *Explicit model uncertainty*: A fully Bayesian procedure has been promoted as a means of weighting alternative models (Morgan et al. 2009), but is essentially reverting to epistemic uncertainty.
- *Qualitative scales expressed as strength of evidence*: This is generally based on hierarchies-of-evidence scales. For example, the US Preventive Services Task Force places its assessments of the net benefit of health-care interventions on a high/moderate/low certainty scale (**http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions**), whereas the UK Crime Reduction Toolkit (**http://whatworks.college.police.uk/toolkit/Pages/Toolkit.aspx**) and the Teaching/Learning Toolkit (**https://educationendowmentfoundation.org.uk/evidence/teaching-learning-toolkit**) both use 5-point scales for strength of evidence. The GRADE (Grades for Recommendation, Assessment, Development and Evaluation) scale (Balshem et al. 2011) is widely used in health applications, placing the certainty of the evidence on a high/moderate/low/very low scale.
- *Acknowledged limitations and their possible impact:* For example, the European Food Safety Authority conducted an exhaustive review of alternative quantitative methods for communicating epistemic uncertainty, but also recommends making qualitative assessments of the impact on a final conclusion from different unmodeled sources of uncertainty (EFSA 2016).
- *Acknowledged ignorance*: Unknown unknowns by definition cannot even be listed, and so strategies that are resilient to surprises are called for. This requires sufficient humility to admit the possibility of being wrong, sometimes known as Cromwell's Law after Oliver Cromwell's celebrated plea to the Church of Scotland: "I beseech you, in the bowels of

Christ, think it possible you may be mistaken" (Carlyle 1871, p. 18). So-called "black swans" (Taleb 2007) need not be unthought-of events: If they are simply more extreme than anything that has occurred before, these can be modeled by an appropriate heavy-tailed distribution.

- *Unacknowledged or meta-ignorance*: This occurs when we do not even consider the possibility of error (Bammer & Smithson 2008), and is to be avoided.

BRIEF CASE STUDIES

Gambling

Gambling is perhaps the archetypal form of risk-taking, and a variety of different ways are used to express the odds, or return on a stake. I shall assume these are all derived from an implicit numerical probability p.

Fractional odds o. This is the traditional form in the United Kingdom. $o = 2/1$ means that for a stake of £1, the punter will get £2 back as well as the original stake of £1. In terms of an underlying probability, $o = (1 - p)/p$, or $p = 1/(1 + o)$. $o = 1/1$ is evens, corresponding to $p = 0.5$.

Moneyline odds m. This format is used in the United States. If m is positive, m represents the profit on a \$100 bet, and if m is negative, it is the stake required to win \$100. So if $p < 1/2$ (lower than evens), $m = 100o = 100(1 - p)/p$, and if $p < 1/2$, $m = -100/o = -100p/(1 - p)$ (somewhat bizarrely for those used to other systems, $m = 100$ and $m = -100$ mean the same thing: evens).

Decimal odds d. This format is used outside the United Kingdom and United States, and is favored by betting exchanges. It is the total return (including the stake, for a successful 1 unit bet). So $d = o + 1 = 1/p$.

A contemporary (at time of writing) example is given in **Table 3**.

This example serves to show that the sophistication of a risk message depends crucially on the context, familiarity, and numeracy of the audience.

Communicating the Potential Benefits and Harms of Pharmaceuticals

There are a wide range of initiatives to provide professionals and public with balanced information about the potential benefits and harms from pharmaceuticals, from regulatory agencies such as the US FDA's Benefit-Risk Assessment Framework (**http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm**) and the effects tables of the EMA (EMA 2014), the BRACE (benefit-risk assessment, communication, and evaluation) project from the pharmaceutical industry (Radawski et al. 2015), and the Drug Facts Box, which in randomized trials has been shown to be generally comprehensible and improve decision-making (Schwartz & Woloshin 2013). The PROTECT (pharmacoepidemiological research on outcomes

Table 3 Odds quoted on at 1 PM on April 29, 2016, in different formats for candidates in the 2016 US presidential election (http://www.oddschecker.com/politics/us-politics/us-presidential-election-2016/winner)

Candidate	Fractional odds	Moneyline odds	Decimal odds	Probability
Hillary Clinton	1/3	−300	1.33	0.75
Donald Trump	7/2	350	4.5	0.22

Table 4 **Extract from Cochrane Summary of Findings Table comparing gabapentin with placebo for neuropathic pain and fibromyalgia (Moore et al. 2014)**

Outcome	Probable outcome with intervention	Probable outcome with placebo	NNT or NNH and/or relative effect (95% CI)	Number of participants	Quality of the evidence (GRADE)	Comments
At least 50% reduction in pain or equivalent	340 in 1,000	210 in 1,000	RR 1.6 (1.3 to 1.9) NNT 8.0 (6.0 to 12)	1,816 (6 studies)	Moderate	Range of doses and dosing regimens pooled to obtain these results, so no guidance regarding efficacy or harm of particular doses
Adverse event withdrawals	110 in 1,000	79 in 1,000	RR 1.4 (1.1 to 1.7) NNH 31 (20 to 66)	4,448 (22 studies)	High	Unlikely new research would change this finding

Abbreviations: CI, confidence interval; GRADE, Working Group grades of evidence; NNH, number needed to harm for adverse event; NNT, number needed to treat; RR, relative risk.

of therapeutics by a European consortium) initiative reviewed alternative methods (Hughes et al. 2016) and visualizations (Hallgreen et al. 2016), recommending close attention to the needs of different audiences.

One of the most developed systems is the Summary of Findings tables provided in systematic reviews by the Cochrane Collaboration (Langendam et al. 2013), an extract of which is shown in **Table 4**.

We note that **Table 4** foregrounds changes in absolute risks, while also providing relative risks and numbers needed to treat or harm: The intervention both increases pain relief and leads to more adverse events. The crucial point is that the potential benefits and harms are communicated in the same format. As well as confidence intervals, a qualitative GRADE assessment is made of the quality of the evidence. Interactive versions of the Summary of Findings Tables are currently being piloted in the DECIDE (developing and evaluating communication to support informed decisions and practice based on evidence) project (**https://isof.epistemonikos.org/#/**).

Climate Change

In such a highly contested area, the framing of risks becomes particularly important, as the example in the sidebar Framing the Risks of Climate Change shows.

The Intergovernmental Panel for Climate Change (IPCC) has expended much effort to standardize their expressions of uncertainty (IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties 2010). **Table 5** shows their agreed definitions of verbal terms for likelihood.

Budescu et al. (2014) tested this proposal on over 13,000 people in 25 countries and 18 languages, and found predictably large variation in interpretation of the verbal terms when the translation table was only available by clicking on a link, with a tendency to regress toward 50%. When the numerical interpretation was provided alongside the verbal term, agreement with the IPCC definitions was better, but still imperfect.

FRAMING THE RISKS OF CLIMATE CHANGE

The United Kingdom's climate impact projections included an assessment of the 90% scenario in the distribution of possible temperature increases. Previously the media had reported such extreme scenarios in terms "temperature rise could be as high as *X*." As a deliberate preemption of this style of coverage, the Department of Environment, Food and Rural Affairs phrased temperature rise as being "very unlikely to be greater than *X*" (Pidgeon & Fischhoff 2011). This change from a negative to a positive frame was successful in deflecting undue attention to the extremes, although a balanced approach would have used both positive and negative frames.

Just as in many health applications, the IPCC has a qualitative scale for the confidence in their analyses, based on quality of evidence and expert agreement. They recommend that the likelihoods in **Table 5** are only used in situations of high or very high confidence. For example, the IPCC (2013, p. 31) reported that "Anthropogenic influences likely contributed to the retreat of the glaciers since the 1960s . . . . Due to a low level of scientific understanding there is low confidence in attributing the causes of the observed loss of mass from the Antarctic ice sheet over the past two decades."

Toxicology and Environmental Exposures

In spite of the prominent public attention given to perceived risks of pesticides, food additives, and other environmental exposures, numerical assessments of risks are almost universally designed for technical audiences. For example, the US EPA's Integrated Risk Information System provides key values for acrylamide (**https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=286**) that include the statements:

- "A reference oral dose (RfD) of 2×10^{-3} mg/kg-day, based on a PoD (point of departure) of 0.053 mg/kg-day and a Composite uncertainty factor of 30, with medium/high confidence." (This means the PoD, based on a transformed lower 95% limit of the estimated dose found to cause adverse events in rats, is then further reduced by a factor of 30 to allow for uncertainties.)
- "A quantitative estimate of carcinogenic risk from oral exposure comprising an oral slope factor of 5×10^{-1} per mg/kg-day." [The oral slope factor is a plausible (upper 95%) upper bound on the estimate of cancer risk per mg/kg-day of lifetime exposure.]

Table 5 Probability intervals corresponding to different verbal terms, as mandated by the Intergovernmental Panel for Climate Change

Term	Likelihood of the outcome (probability)
Virtually certain	99–100%
Extremely likely	95–100%
Very likely	90–100%
Likely	66–100%
More likely than not	50–100%
About as likely as not	33–66%
Unlikely	0–33%
Very unlikely	0–10%
Exceptionally unlikely	0–1%

In another controversial area, the European Food Safety Authority concluded there was no consumer health risk from bisphenol A (BPA) exposure by starting from a potentially hazardous dose of 609 μg/kg-day, then applied a large uncertainty factor of 150 to reach what is termed a tolerable daily intake of 4 μg/kg-day, still hugely above routine exposures (EFSA 2015). Again, the uncertainty factor handles epistemic, crudely quantified uncertainty, whereas overall confidence in the analysis is expressed qualitatively.

There appears to be no attempt at communicating what these terms mean in practice, although there have been recent suggestions for graphics showing how uncertainty leads to the move from PoD to RfD (Beck et al. 2016), and of using the GRADE quality of evidence scale in environmental and occupational health (Morgan et al. 2016). Using examples including acrylamide and BPA, Lofstedt (2013) argued that regulatory agencies need to invest heavily in their communications, with which it would be difficult to disagree.

Security and Intelligence

Public security risk communications are generally phrased in qualitative terms, such as the UK Terrorism Threat Level scale defining SEVERE as meaning "that a terrorist attack is highly likely." Yet when this was level was announced on 22nd January 2010, the Home Secretary felt obliged to say, "This means that a terrorist attack is highly likely, but I should stress that there is no intelligence to suggest than an attack is imminent" (MI5 2010, p. 5).

Marchio (2014) recounted the efforts since the 1960s to bring some quantitative basis to verbal risk communication within the US intelligence services, leading to proposed scales of words of estimative probability similar to those shown in **Table 5**. But current advice from the US National Intelligence Council has backtracked on explicit numerical ratings, and instead they provide a blurred scale between 0 and 1 in which "probably" and "likely" indicate a greater than even chance (**http://nsarchive.gwu.edu/nukevault/ebb270/18.pdf**), and they use "words such as 'we cannot dismiss,' 'we cannot rule out,' and 'we cannot discount' to reflect an unlikely—or even remote—event whose consequences are such it warrants mentioning" (p. 5). They also provide a qualitative assessment of analytic confidence on a high/medium/low scale "based on the scope and quality of information supporting our judgments" (p. 5)—this parallels the approach in both health and climate change.

Explicit numerical risk assessments may, however, be used in internal discussion, as confirmed from this report by Barack Obama on the prospect of the Bin Laden assault: "Some of our intelligence officers thought that it was only a 40 or 30% chance that Bin Laden was in the compound. Others thought that it was as high as 80 or 90%. At the conclusion of a fairly lengthy discussion where everybody gave their assessments I said: this is basically 50–50." (Channel 4 2011). Some have argued that such a diversity of views should have been condensed into a single assessment before being presented to Obama (Friedman & Zeckhauser 2015), but that could have been considered as withholding important information.

Reliability

A bewildering variety of techniques are used in both qualitative and quantitative risk assessment in industrial systems (Melnick & Everitt 2008). For example, Failure Mode and Effects Analysis is a bottom-up approach in which individual components are assessed for probability of failure on, say, a 10-point scale (ASQ 2004), so for example, the US Automotive Industry Action Group (Automot. Ind. Action Group 2009) describe a level-3 probability of failure as low, meaning a less than 1 in 15,000 chance of failure during the scoping interval.

The language of six-sigma may be used as a form of communication of low risks, where deviations from a target are measured in three standard deviation units. So a level-3 failure is also known as "Ppk $>$ 1.33," meaning that a standard normal observation is more than $1.33 \times 3 = 4$ standard deviations from the mean, which occurs with probability 1/15,000. These likelihood assessments, together with severity measures, might then be combined with a fault tree or event tree analysis to give an overall risk level of failure for whole systems.

Weather

Rain is not generally considered a high-risk event, but experience of weather forecasting can be helpful when it comes to greater hazards. For example, the US National Weather Service provides a weather map (**http://www.weather.gov/**) shaded by the chance or probability of precipitation (terms used interchangeably), these assessments being obtained from running an ensemble of models. Probabilities are not so prominently communicated in the United Kingdom, but the UK Meteorological Office does provide precipitation probability (with the lowest possible category being <5%, which says something about the British weather) (**http://www.metoffice.gov.uk/news/in-depth/science-behind-probability-of-precipitation**).

In a valuable review, Stephens et al. (2012) contrast the ambivalent findings regarding frequencies versus percentages in health applications to the clear preference in weather forecasting of chances expressed as a percentage (Morss et al. 2008, Joslyn & Nichols 2009), in spite of the reference class being frequently misunderstood (Gigerenzer et al. 2005).

Natural Hazards

Events such as hurricanes, floods, volcanoes, and earthquakes are potential catastrophes that can affect large numbers of people, and hence there is a focus on mass rather than individual communication. A common feature is the use of geographical hazard maps showing colored regions of different exposure—these are hazards rather than risks, as the impact can clearly be mitigated by precautionary action.

Hurricanes. Research has focused on the US National Hurricane Center's cone of uncertainty, which is defined as follows: "Based on forecasts over the previous 5 years, the entire track of the tropical cyclone can be expected to remain within the cone roughly 60–70% of the time." (**http://www.nhc.noaa.gov/aboutcone.shtml**). This has widespread recognition but is also misinterpreted, in particular there is excessive concentration on the central path (Broad et al. 2007), which can be considered an example of the anchoring heuristic (Kahneman et al. 2006). The cone is presented in a wide variety of formats by different media, sometimes without the central path, and spaghetti plots showing possible paths of the hurricane have also been tried (Stephens et al. 2012). Demuth et al. (2012) emphasized the complexity of communicating hurricane risk, as not only the path but also the storm surge and wind speed are of independent interest.

Floods. A systematic review by Kellens et al. (2013) found little good evidence for the effectiveness of alternative formats. The return period (say 1 in 100 years) is a technical term open to misinterpretation as it suggests a reference class of 100 consecutive years, rather than its correct interpretation as 100 possible next years. Bell & Tobin (2007) found that "1% chance per year" was more effective in conveying uncertainty, but "1 in 100 years" provoked greater concern. The return period is no longer used in the UK Environment Agency flood risk maps (**http://apps.environment-agency.gov.uk/wiyby/37837.aspx**). There appears to be limited understanding and demand for probabilistic systems from civil-protection authorities (Stephens et al. 2012), and a general reluctance to share these with the public, although Pappenberger et al. (2013) proposed a range of appropriate visualizations.

Earthquakes and Volcanoes. There is widespread use of earthquake and volcanic hazard maps; for example, the US Geological Service shows regions with 2% probability of the peak ground acceleration exceeding different levels over a 50-year period (**https://earthquake.usgs.gov/hazards/hazmaps/conterminous/**). In spite of their highly technical definition, the colored maps provide an easily interpretable visualization.

Whereas absolute rather than relative risks are recommended in the health context, this advice can be reversed in acute low-probability, high-impact events that might be mitigated reasonably easily: We all take daily precautions when traveling that make small risks even smaller. For example, Italian earthquake advisors were convicted of manslaughter for issuing unduly reassuring messages to residents of L'Aquila in 2009, a few days before over 300 were killed in a major earthquake (Hall 2011). The advice correctly said that the absolute risk was low, but it has been argued that they should also have emphasized that the *relative* risk was high, thus enabling residents to adopt their own chosen level of precaution (Woo & Marzocchi 2014).

Major Societal Risks and National, Global and Existential Catastrophes

The following examples of low-probability, high-impact events present a particular problem: There is generally limited confidence in the risk assessments because of the inevitable lack of historical evidence on which to base an accurate model, and they can be highly contested, have highly variable outcomes, and cover many orders of magnitude of probability. Communication about these events tends to use technical language and scientific notation, which generally restricts attention to a professional audience skilled in interpreting the measures.

Major Industrial Accidents

A wide variety of risk metrics are used for technical communications involving loss of life and economic damage (Jonkman et al. 2003, Johansen & Rausand 2012). For example, the individual risk per annum (IRPA) is used in safety cases for high-hazard installations provided to the UK Health and Safety Executive, so that the expression $IRPA = 4 \times 10^{-4}$ is an assessment through a quantitative risk analysis that can be translated as meaning each worker has on average a 4 in 10,000 (400 micromorts), or a 1 in 2,500 risk of being killed each year. The use of scientific notation may be fine for a technical audience, but seems almost designed to prevent general comprehension.

F-N curves, originally introduced for the assessment of the risks in the nuclear industry, communicate societal rather than individual risks by plotting the number of lives (x) lost in an accident on the x-axis (log-scale), against the annual probability of exceeding that loss on the y-axis (log-scale), essentially plotting $1 - F(x)$ on a log-log scale. The area under an F-N curve is the expected number of fatalities per year, and a scale aversion index can give additional weight to disasters.

Within the "tolerability of risk" framework (Bouder & Slavin 2007), lines on F-N curves can delineate areas of broadly acceptable and intolerable societal risks. For example, the UK Health and Safety Executive's criteria are shown in **Figure 4**: In particular, an installation with an annual 1 in 5,000 chance of causing 50 deaths is intolerable (Health Saf. Executive 2016). For comparison,

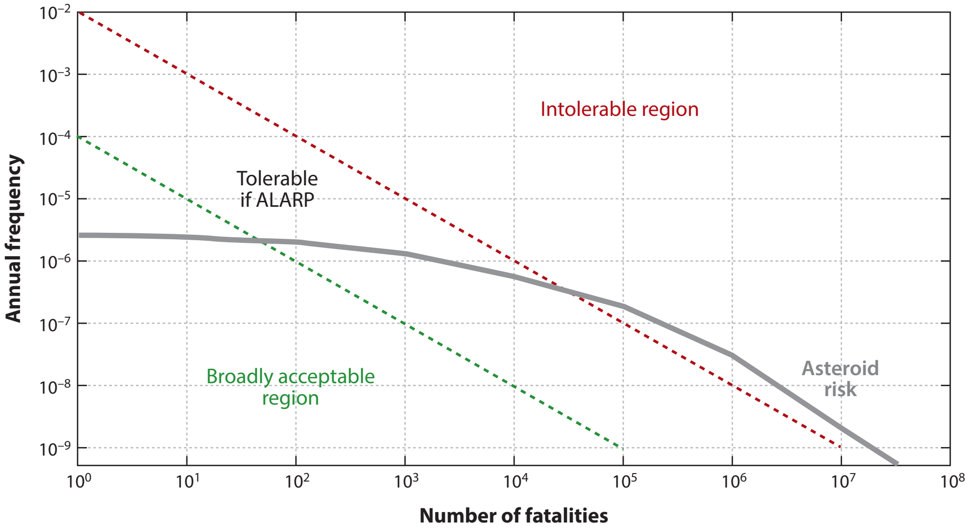

Figure 4

F-N plot showing "broadly acceptable" and "intolerable" regions for industrial installations delineated by the UK Health and Safety Executive: Risks in the area between, though "tolerable," should be made as low as reasonably practicable (ALARP). The gray line approximates the F-N curve of Reinhardt et al. (2016) of the risks from asteroids.

Figure 4 also shows the recently assessed annual global risks from asteroids (Reinhardt et al. 2016) (see the section Global Catastrophes and Existential Risks below).

Catastrophe models are used by insurance companies for predicting financial losses in the event of disasters. A primary form of communication is the exceedance probability curve, which plots various levels of loss against the probability that such a loss will be exceeded, similar to an F-N curve but not always on a log-log scale (and sometimes the axes are reversed). Epistemic uncertainty concerning the strong assumptions in the models is handled through sensitivity analysis of the curve, and nonmodeled uncertainty by appropriate caution in using the outputs (Risk Management Solutions 2012).

National Risks

Many governments carry out risk assessments of events of national importance. For example, the UK National Risk Register (UK Cabinet Off. 2015) assesses rough numerical risks on a mixture of statistical modeling and expertise, and communicates them to an order of magnitude (**Table 6**), although this exercise is sadly given minimal publicity. Security risks are presented without numerical risk assessments, and there is an additional (classified) National Security Risk Assessment (UK Cabinet Off. 2010).

The results of the US Strategic National Risk Assessment "are largely classified and include a comparison of risks for potential incidents in terms of the likelihood (calculated as a frequency—i.e., number of events per year) and consequences of threats and hazards, as well as an analysis of the uncertainty associated with those incidents" (Dep. Homel. Secur. 2011, p. 4). Unfortunately, the unclassified version only lists the events considered, from "biological terrorism attack" to "radiological substance release."

Table 6 UK National Risk Register (UK Cabinet Off. 2015) of nationally important events of defined magnitudes

		Relative likelihood of occurring in the next 5 years				
		Between 1 in 20,000 and 1 in 2,000	Between 1 in 2,000 and 1 in 200	Between 1 in 200 and 1 in 20	Between 1 in 20 and 1 in 2	Greater than 1 in 2
Relative impact score	5				Pandemic influenza	
	4			Coastal flooding Widespread electricity failure		
	3		Major transport accidents Major industrial accidents	Effusive volcanic eruption Emerging infectious disease Inland flooding	Severe space weather Low temperatures and heavy snow Heat waves Poor air quality events	
	2		Disruptive public disorder Severe wildfires	Animal diseases Drought	Explosive volcanic eruption Storms and gales	
	1			Disruptive industrial action		

Major changes since 2012 include the introduction of "widespread electricity failure," whereas "disruptive public disorder" has moved down two likelihood categories (UK Cabinet Off. 2012).

Global Risks

The World Economic Forum produces an annual *Global Risks Report* (World Econ. Forum 2016) in which nearly 750 experts and decision-makers are asked how likely each of 29 risks are to occur globally within the next 10 years, on a scale ranging from 1 (very unlikely) to 7 (very likely)—similar assessments are obtained about impact. At the top of the league table in 2016 was "large-scale involuntary migration" which scored an average of 5.8 in likelihood, up from 16th in 2015, when "interstate conflict" was at the top, suggesting opinions appear to be dominated by recent events.

Global Catastrophes and Existential Risks

The *Annual Report on Global Risks* defines a global catastrophe as killing more than 10% of the world's population (Global Chall. Found. 2016), whereas existential risks threaten human survival itself. Speculative chances of such events are largely based on expert judgment; for example, the median surveyed expert estimated approximately a 19% chance of human extinction by 2100 (although these were participants in a Global Catastrophic Risks conference) (Sandberg & Bostrom 2008).

One of the few global catastrophic risks susceptible to statistical modeling is the threat from asteroids. Popular risk communication uses the Torino scale, ranging from 0 (no likely consequences) to 10 (a collision is certain, capable of causing global climatic catastrophe that may threaten the future of civilization as we know it, whether impacting land or ocean). The NASA Near Earth Object Program currently (as of May 21, 2016) identifies no nonzero threats on the Torino scale, although it provides cumulative impact probabilities in scientific notation (**http://neo.jpl.nasa.gov/risks/**).

Reinhardt et al. (2016) summarized a full risk analysis using an F-N curve reproduced in **Figure 4**, although they used a 100-year rather than an annual frequency scale. They assessed, for example, a 1 in 300 chance that at least one person will be killed from the direct effect of an asteroid over the next 100 years, and a 1 in 300,000 chance of over a million fatalities. If this were an industrial installation, rather than a natural phenomenon, this would be considered an intolerable risk by the UK Health and Safety Executive. Reinhardt et al. (2016) deliberately did not provide an expected number of deaths per year, claiming this is a misleading summary of an extremely skew probability distribution—there is no record of anyone having ever been killed by an asteroid (apart from a cow).

CONCLUSIONS

In spite of the extraordinary range of practices covered in this brief review, and the general lack of firm conclusions, it is still possible to extract some tentative points concerning general recommendations, communicating numbers, and visualizations.

1. General issues when communicating risks based on statistical analysis
   - Be clear about objectives.
   - Segment audience into target groups and identify their needs, beliefs, and skills.
   - Develop, test, and evaluate material with target groups.
   - Build trust by being trustworthy.
   - Use plain language and limit information to only what is necessary.
   - Allow for different levels of interest, knowledge, and numeracy, for example, a top gist level, then numerical information, and then evidence and uncertainty.
   - Have the humility to admit uncertainty.
2. Communicating numerical risks
   - Use absolute risks (but also provide relative risks when dealing with potential catastrophic events).
   - For single unique events, use percent chance if possible, or if necessary, "1 in *X*."
   - When appropriate, express chance as a proportion, a frequency, or a percentage—it is crucial to be clear about the reference class.
   - To avoid framing bias, provide percentages or frequencies both with and without the outcome.
   - Keep the denominator fixed when making comparisons with frequencies, and use an incremental risk format.
   - Be explicit about the time interval.
   - Be aware that comparators can create an emotional response.
   - For more knowledgeable audiences, consider providing quantitative epistemic uncertainty about the numbers and qualitative assessment of confidence in the analysis.
   - More sophisticated metrics can be made for technical audiences, but this only serves to exclude others.
3. Visualizations

   These are derived primarily from Spiegelhalter et al. (2011).
   - Consider a good summary table as a visualization.
   - Use multiple formats, because no single representation suits all members of an audience.
   - Illuminate graphics with words and numbers.
   - Design graphics to allow part-to-whole comparisons on an appropriate scale.
   - Helpful narrative labels are important. Compare magnitudes through tick marks and clearly label comparators and differences.

- Use narratives, images and metaphors that are sufficiently vivid to gain and retain attention, but which do not arouse undue emotion. It is important to be aware of affective responses.
- Assume low numeracy of a general public audience and adopt a less-is-more approach by reducing the need for inferences, making clear and explicit comparisons, and providing optional additional detail.
- Be cautious about interactivity and animations—they may introduce unnecessary complexity.
- Acknowledge the limitations of the information conveyed in its quality and relevance.
- Avoid chart junk, such as three-dimensional bar charts, and obvious manipulation through misleading use of area to represent magnitude.
- Most importantly, assess the needs of the audience, experiment, test, and iterate toward a final design.

The take-home sound bite is simply to be clear about what you are trying to achieve, and check that you are doing it.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful to David Harding and the Winton Foundation for endowing the position I currently hold.

LITERATURE CITED

Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, et al. 2011. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst. Rev.* 3:CD006776

Ancker JS, Senathirajah Y, Kukafka R, Starren JB. 2006. Design features of graphs in health risk communication: a systematic review. *J. Am. Med. Inform. Assoc. JAMIA* 13:608–18

Ancker JS, Weber EU, Kukafka R. 2011. Effect of arrangement of stick figures on estimates of proportion in risk graphics. *Med. Decis. Making* 31:143–50

ASQ (Am. Soc. for Qual.). 2004. *Failure Mode Effects Analysis (FMEA)*. Milwaukee, WI: Am. Soc. for Qual. **http://asq.org/learn-about-quality/process-analysis-tools/overview/fmea.html**

Automot. Ind. Action Group. 2009. *Failure mode and effects analysis ranking table*. **http://elsmar.com/pdf_files/FMEA%20and%20Reliability%20Analysis/AIAG%20FMEA-Ranking-Tables.pdf**

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, et al. 2011. GRADE guidelines: 3. Rating the quality of evidence. *J. Clin. Epidemiol.* 64:401–6

Bammer G, Smithson M, eds. 2008. *Uncertainty and Risk: Multidisciplinary Perspectives*. London: Earthscan

Beck NB, Becker RA, Erraguntla N, Farland WH, Grant RL, et al. 2016. Approaches for describing and communicating overall uncertainty in toxicity characterizations: U.S. Environmental Protection Agency's Integrated Risk Information System (IRIS) as a case study. *Environ. Int.* 89–90:110–28

Bell H, Tobin G. 2007. Efficient and effective? The 100-year flood in the communication and perception of flood risk. *Environ. Hazards* 7:302–11

Bouder F, Slavin D. 2007. *The Tolerability of Risk: A New Framework for Risk Management*. London: Earthscan

Broad K, Leiserowitz A, Weinkle J, Steketee M. 2007. Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bull. Am. Meteorol. Soc.* 88:651–67

Budescu DV, Por H-H, Broomell SB, Smithson M. 2014. The interpretation of IPCC probabilistic statements around the world. *Nat. Clim. Change* 4:508–12

Cancer Research UK. 2015. *Lifetime Risk of Cancer*. London: Cancer Research UK. **http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk**

Carlyle T. 1871. *Oliver Cromwell's Letters and Speeches: With Elucidations*. New York: Scribner, Welford and Co.

Channel 4. 2011. Bin Laden: shoot to kill. *Channel 4 Television Documentary*, first aired Sept. 7

COMEAP (Comm. Med. Eff. Air Pollut.). 2009. *Long-term exposure to air pollution: effect on mortality*. Rep., Comm. Med. Eff. Air Pollut., UK Health Prot. Agency. **https://www.gov.uk/government/publications/comeap-long-term-exposure-to-air-pollution-effect-on-mortality**

Covey J. 2007. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak*. 27:638–54

Cvetkovich G, Lofstedt RE, ed. 2013. *Social Trust and the Management of Risk*. London: Routledge

Demuth JL, Morss RE, Morrow BH, Lazo JK. 2012. Creation and communication of hurricane risk information. *Bull. Am. Meteorol. Soc.* 93:1133–45

Dep. Homel. Secur. 2011. *The Strategic National Risk Assessment in Support of PPD 8: A Comprehensive Risk-Based Approach Toward a Secure and Resilient Nation*. **https://www.dhs.gov/xlibrary/assets/rma-strategic-national-risk-assessment-ppd8.pdf**

Diebold FX, Doherty NA, Herring RJ, eds. 2010. *The Known, the Unknown, and the Unknowable in Financial Risk Management: Measurement and Theory Advancing Practice*. Princeton, NJ: Princeton Univ. Press

Dieckmann NF, Peters E, Gregory R. 2015. At home on the range? Lay interpretations of numerical uncertainty ranges. *Risk Anal*. 35:1281–95

Douglas M, Wildavsky A. 1983. *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. Oakland: Univ. Calif. Press

Ellsberg D. 1961. Risk, ambiguity and the Savage axioms. *Q. J. Econ.* 75:643–69

EFSA (Eur. Food Saf. Auth.). 2016. *Guidance on Uncertainty in EFSA Scientific Assessment*. **https://www.efsa.europa.eu/sites/default/files/160321DraftGDUncertaintyInScientificAssessment.pdf**

EFSA Panel on Food CEF (Eur. Food Saf. Auth. Panel on Food Contact Mater., Enzym., Flavour. and Process. Aids). 2015. Scientific opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. *EFSA J.* 13:3978

EMA (Eur. Med. Agency). 2014. *Benefit-Risk Methodology Project. Update on Work Package 5: Effects Table Pilot (Phase I)*. Eur. Med. Agency, London. **http://www.ema.europa.eu/docs/en_GB/document_library/Report/2014/02/WC500162036.pdf**

EMA (Eur. Med. Agency). 2016. *Section 4.8: Undesirable Effects*. Eur. Med. Agency, London. **http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/01/WC500137021.pdf**

EPA (U.S. Environ. Prot. Agency). 2011. *Vocabulary catalog: Integrated Risk Information System (IRIS) glossary*. **https://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do?details=&glossaryName=IRIS%20Glossary#formTop**

Fagerlin A, Wang C, Ubel PA. 2005. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak*. 25:398–405

Fagerlin A, Zikmund-Fisher BJ, Ubel PA. 2011. Helping Patients Decide: Ten Steps to Better Risk Communication. *JNCI J. Natl. Cancer Inst.* 103:1436–43

Feldman-Stewart D, Brundage MD, Zotov V. 2007. Further insight into the perception of quantitative information: judgments of gist in treatment decisions. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak*. 27:34–43

Few S. 2016. Expressing proportions. *Vis. Bus. Intell. Newsl.*, April/May/June. **http://www.perceptualedge.com/articles/visual_business_intelligence/expressing_proportions.pdf**

Fischhoff B. 1995. Risk perception and communication unplugged: twenty years of Process1. *Risk Anal*. 15:137–45

Fischhoff B. 2006. The science and practice of risk ranking. *Horizons* 10:3. **https://www.cmu.edu/dietrich/sds/docs/fischhoff/SciencePracticeRiskRanking.pdf**

Fischhoff B, Brewer NT, Downs JS, eds. 2011. *Communicating Risks and Benefits: An Evidence-Based User's Guide*. Silver Spring, MD: US Dep. Health Hum. Serv.

Fischhoff B, Davis AL. 2014. Communicating scientific uncertainty. *PNAS* 111:13664–71
Friedman JA, Zeckhauser R. 2015. Handling and mishandling estimative probability: likelihood, confidence, and the search for Bin Laden. *Intell. Natl. Secur.* 30:1–23
Gaissmaier W, Wegwarth O, Skopec D, Müller A-S, Broschinski S, Politi MC. 2012. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. *Health Psychol.* 31:286–96
Galesic M, Garcia-Retamero R. 2010. Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Arch. Intern. Med.* 170:462–68
Galesic M, Garcia-Retamero R, Gigerenzer G. 2009. Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc.* 28:210–16
Galesic M, Kause A, Gaissmaier W. 2016. A Sampling Framework for Uncertainty in Individual Environmental Decisions. *Top. Cogn. Sci.* 8:242–58
Garcia-Retamero R, Galesic M, Gigerenzer G. 2010. Do icon arrays help reduce denominator neglect? *Med. Decis. Mak.* 30:672–84
Geosci. Aust. 2014. *What is Risk?* Geosci. Aust., Symonston, Aust. **http://www.ga.gov.au/scientific-topics/hazards/risk-impact/risk**
Gigerenzer G. 2014. Should patients listen to how doctors frame messages? *BMJ* 349:g7091
Gigerenzer G, Galesic M. 2012. Why do single event probabilities confuse patients? *BMJ* 344:e245
Gigerenzer G, Hertwig R, van den Broek E, Fasolo B, Katsikopoulos KV. 2005. "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Anal.* 25:623–29
Global Chall. Found. 2016. *Global Catastrophic Risks*. Global Chall. Found., Stockholm. **http://globalchallenges.org/annual-report-on-global-risks**
Goldin R. 2015. Death by bacon: Did the news get to the meat of the matter? *STATS*, Nov. 11. **http://www.stats.org/death-by-bacon-did-the-news-get-to-the-meat-of-the-matter/**
Hall SS. 2011. Scientists on trial: at fault? *Nat. News* 477:264–69
Hallgreen CE, Mt-Isa S, Lieftucht A, Phillips LD, Hughes D, et al. 2016. Literature review of visual representation of the results of benefit–risk assessments of medicinal products. *Pharmacoepidemiol. Drug Saf.* 25:238–50
Han PKJ, Klein WMP, Lehman T, Killam B, Massett H, Freedman AN. 2011. Communication of uncertainty regarding individualized cancer risk estimates. *Med. Decis. Mak.* 31:354–66
Harmsen CG, Kristiansen IS, Larsen PV, Nexoe J, Stovring H, et al. 2014. Communicating risk using absolute risk reduction or prolongation of life formats: cluster-randomised trial in general practice. *Br. J. Gen. Pract.* 64:e199–207
Health Saf. Executive. 2016. *HID's Approach to ALARP Decisions—SPC/Permissioning/39*. Health Saf. Executive, Bootle, UK. **http://www.hse.gov.uk/foi/internalops/hid_circs/permissioning/spc_perm_39.htm**
Hildon Z, Allwood D, Black N. 2012. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. *Int. J. Qual. Health Care* 24:55–64
Hughes D, Waddingham E, Mt-Isa S, Goginsky A, Chan E, et al. 2016. Recommendations for benefit-risk assessment methodologies and visual representations. *Pharmacoepidemiol. Drug Saf.* 25:251–62
IPCC (Intergov. Panel Clim. Change). 2013. Summary for policymakers. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press. **https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_SPM_FINAL.pdf**
IPCC (Intergov. Panel Clim. Change) Cross-Working Group Meeting on Consistent Treatment of Uncertainties. 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *IPCC Cross-Work. Group Meet. Consistent Treat. Uncertain., Jasper Ridge, CA, USA, 6–7 July*.
ISO (Int. Stan. Organ.). 2009. *ISO 31000—Risk Management*. Geneva: ISO. **http://www.iso.org/iso/home/standards/iso31000.htm**
Johansen IL, Rausand M. 2012. Risk metrics: Interpretation and choice. *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, pp. 1914–18. Piscataway, NJ: IEEE
Jonkman SN, van Gelder PHAJM, Vrijling JK. 2003. An overview of quantitative risk measures for loss of life and economic damage. *J. Hazard. Mater.* 99:1–30

Joslyn SL, Nichols RM. 2009. Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorol. Appl.* 16:309–14

Kahan DM. 2015. What is the "science of science communication"? *J. Sci. Comm.* 14:1–10

Kahneman D, Slovic P, Tversky A. 2006. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge Univ. Press

Kellens W, Terpstra T, De Maeyer P. 2013. Perception and communication of flood risks: a systematic review of empirical research. *Risk Anal.* 33:24–49

Keller C, Siegrist M, Visschers V. 2009. Effect of risk ladder format on risk perception in high- and low-numerate individuals. *Risk Anal.* 29:1255–64

Kemp R. 2016. Why are only 1 in 100 men taking up shared parental leave? *The Telegraph*, Apr. 5. **http://www.telegraph.co.uk/men/the-filter/why-are-only-1-in-100-men-taking-up-shared-parental-leave/**

Kirby J. 2015. Bacon, ham and sausages have the same cancer risk as cigarettes warn experts. *Daily Record*, Oct. 23. **http://www.dailyrecord.co.uk/news/health/bacon-ham-sausages-same-cancer-6689230#ebyBMgxErTb9Jf7i.97**

Knapp P. 2004. Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Qual. Saf. Health Care* 13:176–80

Knapp P, Gardner PH, Woolf E. 2016. Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations. *Health Expect.* 19:264–74

Knight F. 1921. *Risk, Uncertainty and Profit*. Boston: Hart, Schaffner & Marx

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.* 36:1599–614

Lachlan K, Spence PR. 2010. Communicating risks: examining hazard and outrage in multiple contexts. *Risk Anal.* 30:1872–86

Langendam MW, Akl EA, Dahm P, Glasziou P, Guyatt G, Schünemann HJ. 2013. Assessing and presenting summaries of evidence in Cochrane Reviews. *Syst. Rev.* 2:81

Leiss W. 1996. Three phases in the evolution of risk communication practice. *Ann. Am. Acad. Pol. Soc. Sci.* 545:85–94

Lempert RJ, Collins MT. 2007. Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches. *Risk Anal.* 27:1009–26

Lofstedt R. 2013. Communicating food risks in an era of growing public distrust: three case studies: perspective. *Risk Anal.* 33:192–202

Lopez-Gonzalez AA, Aguilo A, Frontera M, Bennasar-Veny M, Campos I, et al. 2015. Effectiveness of the Heart Age tool for improving modifiable cardiovascular risk factors in a Southern European population: a randomized trial. *Eur. J. Prev. Cardiol.* 22:389–96

Marchio J. 2014. "If the weatherman can...": the intelligence community's struggle to express analytic uncertainty in the 1970s. *Stud. Intell.* 58:31–42

Melnick EL, Everitt BS. 2008. *Encyclopedia of Quantitative Risk Analysis and Assessment*. New York: John Wiley & Sons

MI5. 2010. *MI5 Terrorism Threat Level*. **https://www.mi5.gov.uk/threat-levels**

Moore RA, Wiffen PJ, Derry S, Toelle T, Rice ASC. 2014. Gabapentin for chronic neuropathic pain and fibromyalgia in adults. *Cochrane Database Syst. Rev.* (3): CD007938

Morgan G, Dowlatabadi H, Henrion M, Keith D, Lempert R, et al. 2009. *Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decision Making*. Silver Spring, MD: Natl. Ocean. Atmos. Organ.

Morgan MG, Fischhoff B, Bostrom A, Atman CJ. 2001. *Risk Communication: A Mental Models Approach*. Cambridge, UK: Cambridge Univ. Press. 1st ed.

Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, et al. 2016. GRADE: assessing the quality of evidence in environmental and occupational health. *Environ. Int.* 92–93:611–16

Morss R, Demuth J, Lazo J. 2008. Communicating uncertainty in weather forecasts: a survey of the U.S. public. *Weather Forecast.* 23:974–91

NRC (Natl. Res. Counc.). 1989. *Improving Risk Communication*. Washington, D.C.: Natl. Acad. Press

Paling J. 2003. Strategies to help patients understand risks. *BMJ* 327:745–48

Pappenberger F, Stephens E, Thielen J, Salamon P, Demeritt D, et al. 2013. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrol. Process.* 27:132–46

Peters E. 2008. Numeracy and the Perception and Communication of Risk. *Ann. N. Y. Acad. Sci.* 1128:1–7

Peters E, Dieckmann NF, Västfjäll D, K C, Slovic P, Hibbard JH. 2009. Bringing meaning to numbers: The impact of evaluative categories on decisions. *J. Exp. Psychol. Appl.* 15:213–27

Peters E, Hart PS, Fraenkel L. 2011. Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Med. Decis. Mak*. 31:432–36

Pidgeon N, Fischhoff B. 2011. The role of social and decision sciences in communicating uncertain climate risks. *Nat. Clim. Change* 1:35–41

Pighin S, Savadori L, Barilli E, Galbiati S, Smid M, et al. 2015. Communicating Down syndrome risk according to maternal age: "1-in-X" effect on perceived risk. *Prenat. Diagn.* 35:777–82

Radawski C, Morrato E, Hornbuckle K, Bahri P, Smith M, et al. 2015. Benefit-risk assessment, communication, and evaluation (BRACE) throughout the life cycle of therapeutic products: overall perspective and role of the pharmacoepidemiologist. *Pharmacoepidemiol. Drug Saf.* 24:1233–40

Rakow T, Wright RJ, Bull C, Spiegelhalter DJ. 2012. Simple and multistate survival curves: Can people learn to use them? *Med. Decis. Mak*. 32:792–804

Rakow T, Wright RJ, Spiegelhalter DJ, Bull C. 2015. The pros and cons of funnel plots as an aid to risk communication and patient decision making. *Br. J. Psychol.* 106:327–48

Reinhardt JC, Chen X, Liu W, Manchev P, Paté-Cornell ME. 2016. Asteroid risk assessment: a probabilistic approach. *Risk Anal*. 36:244–61

Risk Management Solutions. 2012. *Model use and uncertainty*. **http://www.casact.org/community/affiliates/camar/1012/uncertainty.pdf**

Ropeik D. 2010. *How Risky Is It, Really? Why Our Fears Don't Always Match the Facts*. New York: McGraw-Hill

Rumsfeld D. 2002. *Defense.gov News Transcript.* **http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID = 2636**

Sandberg A, Bostrom N. 2008. *Global catastrophic risks survey*. Tech. Rep. 2008–1, Future Humanity Inst., Univ. Oxford

Sandman PM, Miller PM, Johnson BB, Weinstein ND. 1993. Agency communication, community outrage, and perception of risk: three simulation experiments. *Risk Anal*. 13:585–98

Sandman PM, Weinstein ND, Miller P. 1994. High risk or low: how location on a "risk ladder" affects perceived risk. *Risk Anal*. 14:35–45

Savage LJ. 1951. The Theory of Statistical Decision. *J. Am. Stat. Assoc.* 46:55–67

Savage LJ. 1972. *The Foundations of Statistics*. New York: Dover

Scheer D, Benighaus C, Benighaus L, Renn O, Gold S, et al. 2014. The distinction between risk and hazard: understanding and use in stakeholder communication. *Risk Anal*. 34:1270–85

Schwartz LM, Woloshin S. 2013. The drug facts box: improving the communication of prescription drug information. *PNAS* 110:14069–74

Sirota M, Juanchich M, Kostopoulou O, Hanak R. 2014. Decisive evidence on a smaller-than-you-think phenomenon: revisiting the "1-in-x" effect on subjective medical probabilities. *Med. Decis. Mak*. 34:419–29

Skarlatidou A, Cheng T, Haklay M. 2012. What do lay people want to know about the disposal of nuclear waste? A mental model approach to the design and development of an online risk communication. *Risk Anal*. 32:1496–511

Slovic P. 1987. Perception of risk. *Science* 236:280–85

Slovic P. 2000. *The Perception of Risk*. England: Earthscan

Slovic P, Peters E. 2006. Risk perception and affect. *Curr. Dir. Psychol. Sci.* 15:322–25

Society of Thoracic Surgeons. 2015. Adult cardiac surgery database executive summary. Chicago: Soc. Thorac. Surg. **http://www.sts.org/sites/default/files/documents/2015Harvest4_ExecutiveSummary.pdf**

Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexøe J, Nielsen JB. 2008. Laypersons' understanding of relative risk reductions: Randomised cross-sectional study. *BMC Med. Inform. Decis. Mak*. 8:31

Spiegelhalter DJ. 2002. Mortality and volume of cases in paediatric cardiac surgery: retrospective study based on routinely collected data. *BMJ* 324:261–61

Spiegelhalter DJ. 2012. Using speed of ageing and "microlives" to communicate the effects of lifetime habits and environment. *BMJ* 345:e8223

Spiegelhalter DJ. 2014. The power of the MicroMort. *BJOG Int. J. Obstet. Gynaecol.* 121:662–63

Spiegelhalter DJ, Pearson M, Short I. 2011. Visualizing uncertainty about the future. *Science* 333:1393–400

Spiegelhalter DJ, Riesch H. 2011. Don't know, can't know: embracing deeper uncertainties when analysing risks. *Philos. Trans. R. Soc. A* 369:4730–50

Stephens EM, Edwards TL, Demeritt D. 2012. Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction. *Wiley Interdiscip. Rev. Clim. Change* 3:409–26

Stirling A. 2007. Risk, precaution and science: towards a more constructive policy debate. Talking point on the precautionary principle. *EMBO Rep*. 8:309–15

Stone ER, Sieck WR, Bull BE, Yates JF, Parks SC, Rush CJ. 2003. Foreground:background salience: explaining the effects of graphical displays on risk avoidance. *Organ. Behav. Hum. Decis. Process.* 90:19–36

Tait AR, Voepel-Lewis T, Zikmund-Fisher BJ, Fagerlin A. 2010. Presenting research risks and benefits to parents: Does format matter? *Anesth. Analg.* 111:718–23

Taleb N. 2007. *The Black Swan: the Impact of the Highly Improbable*. New York: Random House. 1st ed.

Trevena LJ, Barratt A, Butow P, Caldwell P. 2006. A systematic review on communicating with patients about evidence. *J. Eval. Clin. Pract.* 12:13–23

Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, et al. 2013. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med. Inform. Decis. Mak*. 13(Suppl 2):S7

UK Cabinet Off. 2010. *Fact Sheet2: National Security Risk Assessment*. London: Cabinet Off. **https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62484/Factsheet2-National-Security-Risk-Assessment.pdf**

UK Cabinet Off. 2011. *Communicating Risk Guidance*. London: Cabinet Off. **https://www.gov.uk/government/publications/communicating-risk-guidance**

UK Cabinet Off. 2012. *National Risk Register for Civil Emergencies—2012 Update*. London: Cabinet Off. **https://www.gov.uk/government/publications/national-risk-register-for-civil-emergencies-2012-update**

UK Cabinet Off. 2015. *National Risk Register for Civil Emergencies—2015 Edition*. London: Cabinet Off. **https://www.gov.uk/government/publications/national-risk-register-for-civil-emergencies-2015-edition**

van der Linden S. 2016. A conceptual critique of the cultural cognition thesis. *Sci. Commun.* 38:128–38

van Dyk DA. 2014. The role of statistics in the discovery of a Higgs boson. *Annu. Rev. Stat. Appl.* 1:41–59

Visschers VHM, Meertens RM, Passchier WWF, de Vries NNK. 2009. Probability information in risk communication: a review of the research literature. *Risk Anal*. 29:267–87

WHO (World Health Organ.). 2016. *WHO global urban ambient air pollution database*. World Health Organ., updated 2016, retrieved May 24, 2016. **http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/**

Woo G, Marzocchi W. 2014. Operational earthquake forecasting and decision-making. In *Early Warning for Geological Disasters*, ed. F Wenzel, J Zschau, pp. 353–67. Berlin: Springer

World Econ. Forum. 2016. *The Global Risks Report 2016*, *11th Edition*. Geneva: World Econ. Forum. **https://www.weforum.org/reports/the-global-risks-report-2016/**

Wynne B. 1991. Knowledges in Context. *Sci. Technol. Hum.* 16:111–21

Yamagishi K. 1997. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Appl. Cogn. Psychol.* 11:495–506

Zikmund-Fisher BJ. 2014. Continued use of 1-in-x risk communications is a systemic problem. *Med. Decis. Making* 34:412–13

Zikmund-Fisher BJ, Dickson M, Witteman HO. 2011. Cool but counterproductive: interactive, web-based risk communications can backfire. *J. Med. Internet Res.* 13:e60

Zikmund-Fisher BJ, Fagerlin A, Ubel PA. 2010. A demonstration of "'less can be more'" in risk graphics. *Med. Decis. Mak*. 30:661–71

Zikmund-Fisher BJ, Ubel PA, Smith DM, Derry HA, McClure JB, et al. 2008. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Educ. Couns.* 73:209–14

Zikmund-Fisher BJ, Witteman HO, Fuhrel-Forbis A, Exe NL, Kahn VC, Dickson M. 2012. Animated graphics for comparing two risks: a cautionary tale. *J. Med. Internet Res.* 14:e106

Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, et al. 2014. Evidence-based risk communication: a systematic review. *Ann. Intern. Med.* 161:270–80

Annual Review of Statistics and Its Application

Volume 4, 2017

Contents

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics