Article

Policy Insights from the Behavioral and Brain Sciences
2019, Vol. 6(1) 29–37

DOI: 10.1177/2372732218813377
journals.sagepub.com/home/bbs
SAGE

Helping People Know Whether Measurements Have Good or Bad Implications: Increasing the Evaluability of Health and Science Data Communications

Brian J. Zikmund-Fisher[1]

Abstract

Patients, policy makers, and the public have access to many types of health and scientific data relevant to various individual and societal decisions. Yet, these audiences often struggle with the meaning and the potential usefulness of those data, so they may choose not to engage with the data at all. Scientific and health data are generally difficult to interpret, yet presentations often implicitly assume that the recipient has the necessary contextual knowledge to identify the data's meaning. To address this problem, designers of data communications should go beyond considering audience characteristics (e.g., numeracy) and focus more on increasing information evaluability (a concept from the judgment and decision-making literatures). The challenge is understanding which data characteristics guide people's ability to extract meaning from data in a given situation. Prioritizing use-relevant contextual information (e.g., by defining action thresholds, comparison standards, meaningful categories, and/or significant differences) is the single best thing experts can do to improve data communication effectiveness. Doing so increases the chances that the patient, public, or policy maker audience does not just know what their numbers are but also what they mean.

Keywords

science communication, health communication, data communication

Tweet

Make health/science data communications more meaningful for public or policy audiences: Provide use-relevant context to increase evaluability.

Key Points

- Many health and science data are both unfamiliar to patients, policy makers, and the public and difficult for them to understand.
- Contextual information determines whether data are easy or difficult for people to make sense of and use in decision making.
- The single best thing that communicators of health and science data can do to improve their communication effectiveness is to increase information evaluability by providing use-relevant contextual information.
- A data-evaluability worksheet can help communicators of health and science data to determine what types of contextual information will be most helpful to their target audiences.
- Anyone communicating scientific or health measurements to nonexpert audiences should provide use-relevant contextual information to enable recipients to determine the data's good or bad implications the way an expert does.

Introduction

In 1999, a highly educated and numerate patient was in the hospital recovering from a very risky bone marrow transplant procedure. Each day, he had tubes of blood drawn for various laboratory tests, and each day he checked his records to see that day's results. One day, he noted that his bilirubin, a measure of liver function, was higher than it had been before. Concerned that this indicated some emerging problem, he asked the attending physician about the result. Her answer was simple: "Don't worry about that. I'll tell you when to worry." The patient was unhappy with that response.

The above story is true, and I was that patient. I recount this story because it illustrates four critical foundations of

[1]University of Michigan, Ann Arbor, USA

Corresponding Author:
Brian J. Zikmund-Fisher, Department of Health Behavior and Health Education, University of Michigan, 1415 Washington Heights, Ann Arbor MI, 48109-2029, USA.
Email: bzikmund@umich.edu

this article: First, health and scientific data, such as laboratory test results, are increasingly available to nonprofessional audiences. In fact, today, most patients in the United States can directly access their laboratory test results online through patient portals to electronic health record systems (American Hospital Association, 2016). Second, even though I was a PhD student in decision psychology and behavioral economics at the time of this story, I still was unable to determine the meaning of these accessible test result data. Third, the expert in this story, my doctor, knew immediately whether the result was good or bad, significant or irrelevant, based on her training and expertise. Fourth, my reaction to my doctor's statement illustrates the fact that having unusable data is frustrating.

Patients, policy makers, and the public need scientific data for health and environmental decision making. But, both my personal experiences and 20 years of conducting research on design of health data communications lead me to believe the core problem is *not* finding or gaining access to the best data. Access to data is, of course, important. The biggest challenge in enabling data usage, however, involves making the available data actually meaningful to nonexpert users by recognizing what specific data characteristics guide meaningful interpretation and providing relevant reference standards.

This argument relies on a simple concept from cognitive science: information evaluability (Hsee, 1996). Some attributes of products are easy to evaluate independently, that is, we immediately know whether they are good or bad. Other attributes are hard to evaluate without explicit comparison standards. Because of differences in attribute evaluability, decision makers tend to rely upon easily understandable data and ignore the more difficult-to-interpret data. This may cause preference reversals between joint and separate consideration of alternatives (Hsee, 1996). That is, options that are comparatively good on easy-to-evaluate attributes but poor on hard-to-evaluate attributes might be rated high when considered in isolation (separate evaluation). But the same options might be rejected in side-by-side comparisons (joint evaluation) with options that are stronger on the hard-to-evaluate attribute. For example, a music player (e.g., iPod) with poor sound distortion levels (a hard-to-evaluate attribute) but high song capacity (an easy-to-evaluate attribute) is rated highly by itself but much lower when compared against another player with better distortion levels. Thus, evaluability issues could explain why people sometimes choose options that are clearly less good than other options (Hsee, 1998).

Evaluability theory has evolved over time (Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2010), with later papers noting that evaluability changes with different types of reference information. For example, knowing the best possible and worst possible values of an attribute is fundamentally different than knowing its average value, even though both improve the evaluability of a given score. General evaluability theory argues that evaluability depends on three factors (Hsee & Zhang, 2010): (a) Evaluability of a value depends on whether it is considered by itself or jointly, the user's prior knowledge, and whether the attribute is inherently easy to interpret or not. Inherently evaluable values are those for which "human beings have an innate and stable physiological or psychological 'scale' (reference system)" (p. 345) that aids in interpretation. (b) All three factors have to be poor in order for people to have no sensitivity to variations. (c) If a given value is easy for communicators to evaluate, they will tend to overestimate other people's ability to interpret and use it.

While evaluability research appears primarily in the judgment and decision psychology and the marketing literatures, poor evaluability is a major barrier to the public's ability to understand and use health and science data. Most people are not scientists or doctors. They do not see blood counts, chemical concentration data, global temperature averages, or airborne particulate rates on a daily basis. They have no formal training and may have little experience in interpreting these data. Moreover, these data represent constructs that for the most part are intangible to humans. In short, these data are inherently difficult to understand, most people lack relevant knowledge, and the data are often presented to the public in isolation. Is it any surprise that patients, the public, and policy makers struggle to make use of them? Note also that evaluability theory suggests that medical professionals and scientists will themselves struggle to recognize how hard it is for these audiences to derive meaning from their data.

One critical question that anyone intending to communicate unfamiliar data should ask themselves is, how should communicators identify and select the right contextual information to maximize usability of the data by their target audience? In this article, I first identify the evaluability-related barriers to interpretation of various types of health and science data. I then describe an approach to designing data communications that focuses on understanding which data characteristics are central to data evaluability in a given situation. Finally, I frame consideration of information evaluability as an example of the larger need for message prioritization in scientific communication.

Evaluability Issues in Health and Science Data Communications

Evaluability challenges exist in many health and science data communication domains. This section of the article discusses several speculative examples.

Medical Test Results

Clinical laboratory test results exemplify hard-to-evaluate data for most patients: Results usually appear in isolation, many patients have little to no prior knowledge about good or bad values, and the test value is inherently hard to evaluate because units are unfamiliar and tests vary in whether higher values represent better or worse outcomes. While current test result communications include a "standard range," that one reference standard is insufficient to clarify just how bad out-of-range values might be. Many patients also conduct regular biomarker testing at home (e.g., blood pressure tests, blood glucose testing). However, visual displays—especially those that improve evaluability through explicit categorization of values, action thresholds, or target ranges (Figure 1)—can improve people's understanding and, more importantly, their sensitivity to variations in test values (Scherer et al., 2018; Zikmund-Fisher et al., 2018; Zikmund-Fisher et al., 2017).

Environmental Exposure Levels

Understanding environmental exposure data is similarly critical for policy makers and the public to make evidence-informed decisions to minimize health risks related to contaminants such as radon, arsenic, lead, ozone, and airborne particulates. Yet, these results are unfamiliar and use difficult-to-interpret units (e.g., picocuries per liter of air). Although direction of improvement is generally clear (less contaminant is better), most people have no idea how to interpret a single observation without comparison to a reference standard (e.g., the U.S. Environmental Protection Agency's radon action threshold of 4 pCi/L). This lack of evaluability contributes to people's desire for zero exposure and, hence, zero risk (Baron, Gowda, & Kunreuther, 1993; Ritov, Baron, & Hershey, 1993), because the zero point is the only easy-to-understand reference point, even if it is unachievable in practice.

Climate Observations

Evaluability issues arise regarding many common measures of climate change (e.g., carbon dioxide concentrations, global average temperatures). For example, people have an intuitive sense of how temperatures are likely to vary in our home environment and know that variations of a degree or two in daily temperatures are generally unimportant. Yet, measures of global average temperature are conceptually different, and a change in global average temperature levels by only a few degrees represents a significant shift in global climate (National Research Council, 2010). Understanding global temperature values is hence an evaluability problem: Even though observations look similar to local daily temperature values, users have to understand the radically narrower range of variation and different reference standards to accurately interpret these data.

Evaluations of Quality of Life or Environment

Evaluability is also central whenever the public is asked to value different health states or different environmental outcomes (e.g., as an input into cost-effectiveness analyses, risk management decisions, or environmental damage evaluations). Evaluations of different levels of disability or health states, for example, depend on whether the questions are asked in separate or joint evaluation and on what comparison standards are provided, although the exact mechanism of the effect remains in dispute (Lacey, Loewenstein, & Ubel, 2011; Pinto-Prades, Robles-Zurita, Sánchez-Martínez, Abellán-Perpiñán, & Martínez-Pérez, 2017).

Risk Estimates

Is a 28% success rate good or bad? Without more information, it is impossible to know, and hence people tend to ignore single risk statistics in decision making (Zikmund-Fisher, 2013; Zikmund-Fisher, Fagerlin, & Ubel, 2004). Risk perceptions change substantially depending on the presence of comparative information, such as the risk of the average person (Fagerlin, Zikmund-Fisher, & Ubel, 2007). People's perceptions of risk reduction from drug therapies change when alternatives are presented in joint versus separate evaluation (Gyrd-Hansen et al., 2011). To increase evaluability, patient decision aids often facilitate side-by-side comparisons of treatment options on relevant success rates or side effect risks. Similar side-by-side comparison tools are also common on shopping websites.

Disease Case Counts

In epidemics or natural disaster situations, public health officials often communicate case counts or mortality data to policy makers and the public. But, how concerned should one be about 10 cases of a disease, or 100, or 1,000? While fewer cases are always better, judgments of whether a disease should be a priority require that users know whether 100 cases represent a major outbreak (e.g., in an Ebola outbreak) or is consistent with prevailing levels of endemic disease (e.g., in tracking cases of Lyme disease). Similarly, case counts during the 2009 H1N1 influenza outbreak were best understood in comparison with average seasonal influenza outcomes rather than in isolation.

Personal Fitness Devices

Personal activity trackers provide access to data about step counts, heart rates, minutes of activity, and so on. While the direction of improvement for these measures is clear (more exercise is better), evaluating a particular step count (e.g., 7,000 steps) is difficult. While 7,000 steps are below the commonly discussed standard of 10,000 steps per day, it may compare favorably with past activity levels, the performance of others in a workplace walking group, or a personalized target goal. Highlighting different comparison standards plausibly alters whether people feel motivated or demotivated by their step counts.

Table with Standard Range:

Test	Your Result	Standard Range	Units
Platelet Count (PLT)	135	150-400	x10^9/L

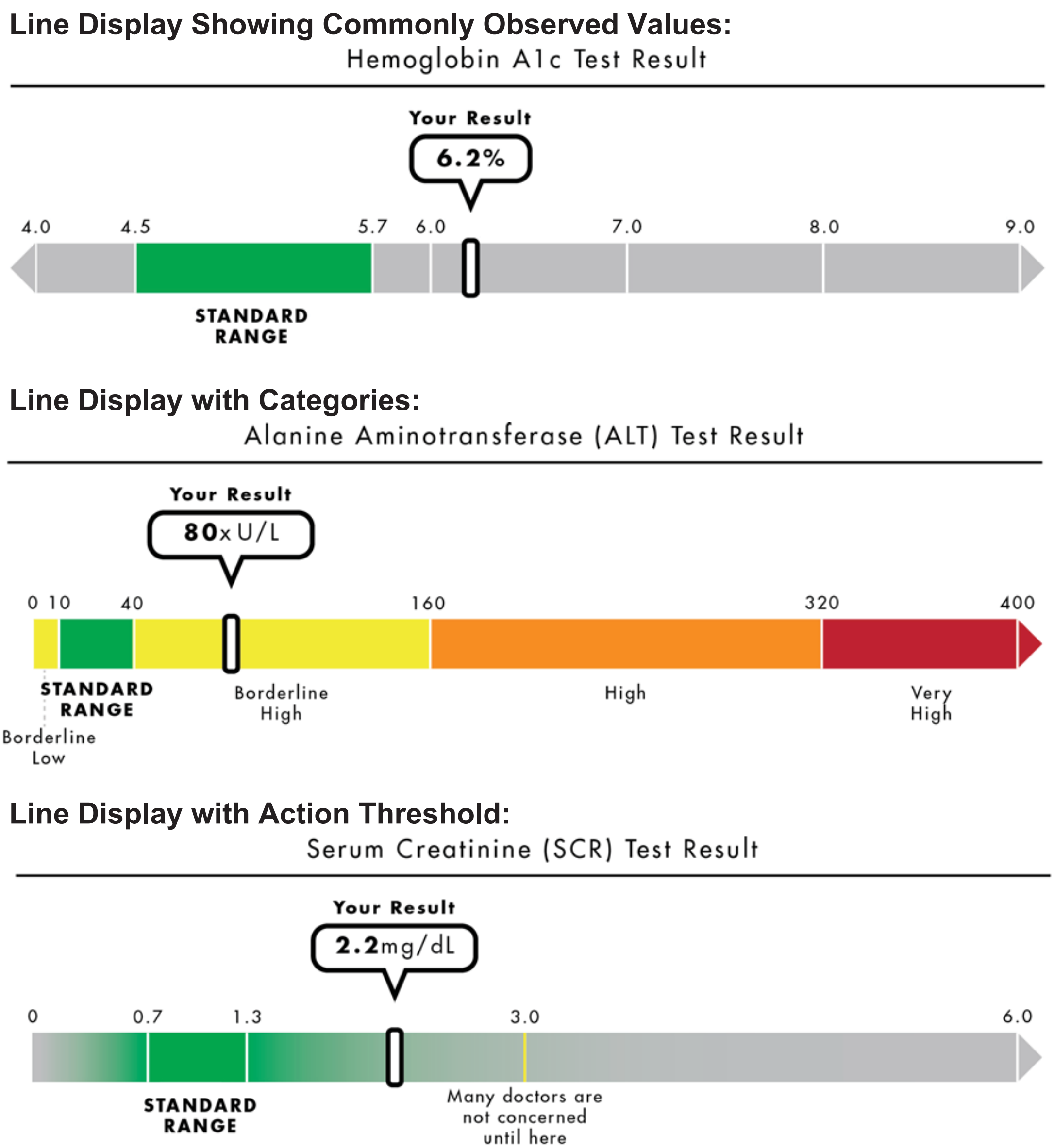

Figure 1. Laboratory test result displays that highlight various data characteristics and reference standards that affect evaluability.
Source. Scherer et al. (2018), Zikmund-Fisher et al. (2018), Zikmund-Fisher et al. (2017).

Food Labels

Food labels now commonly include calorie counts and amounts of sodium, fat, or fiber, which are all difficult-to-evaluate types of data. Regulations requiring restaurants to post calorie counts on menus are an attempt to make calorie information not just accessible but more evaluable (because one can compare different items). However, evaluability challenges exist: Because no clear standards define how many calories an entrée or dish should have and people have a poor sense of how many calories they should consume daily (Krukowski, Harvey-Berino, Kolodinsky, Narsana, & DeSisto, 2006), an item will seem better (but not actually be healthier) when presented next to high calorie alternatives. In grocery stores, evaluative label systems (e.g., color-coded labels for a particular fat or calorie attribute) could benefit consumers, however. These approaches increase the meaningfulness of nutrition information by enabling consumers to respond to a simple categorization system rather than have to interpret the underlying health data (Newman, Howlett, & Burton, 2014).

Health Insurance Plan Attributes

Selecting a health insurance plan requires trading off cost (an easy-to-evaluate attribute) versus a variety of better or worse coverage benefits, all of which are unfamiliar and difficult to evaluate (Hibbard & Peters, 2003). Some attributes, such as hospital or health system quality, are inherently comparative (i.e., one hospital vs. another). Others, such as number of covered mental health visits, should ideally be evaluated on an absolute basis. Transforming quality or coverage measures into common scales increases evaluability and hence usability of this information. The categorization of health plans in health insurance marketplaces (e.g., into Bronze, Silver, and Gold plans) is another (very coarse) approach to increasing evaluability in this context.

Identifying Critical Context: The Data Evaluability Characteristics Worksheet

Both evaluability theory and the examples above suggest that the key to designing effective communications of hard-to-evaluate data is *identifying the data characteristics* that people need to use to evaluate an unfamiliar value in productive ways. Once having identified these critical comparison standards, communicators can design their message accordingly. The more accurate the selection process, the more useful the message will be.

Unfortunately, most health professionals and scientists do not consider the evaluability characteristics of their data. The reason is obvious: Experts already know the range of possible values, they have expectations about good versus poor values, they know population norms, and so on. As they have high knowledge, they do not need to think about these questions to use the data.

The solution is to increase conscious consideration of data characteristics by communicators. In teaching, I demonstrate the importance of this step by having students complete a Data Evaluability Characteristics Worksheet (Table 1). The worksheet includes questions and concepts derived both from research and from materials developed as part of a consensus report on whether and how to return individual-specific test results to research participants (National Academies of Sciences, Engineering, and Medicine, 2018).

The worksheet starts with basic information: What are the units (e.g., % vs. parts per trillion)? What is the theoretically possible range of values (e.g., 0-100, or 0 to infinity)? Which direction is better (more, less, or optimal range)? This last question is deceptively important. For example, having higher levels of low-density lipoprotein (LDL, or "bad" cholesterol) increases cardiovascular risk, but higher levels of high-density lipoprotein (HDL or "good" cholesterol) does the reverse. Communicators should never assume that people know whether more is good or bad.

Table 1. Data Evaluability Characteristics Worksheet.

Background:

- What is the technical term used to describe this test or measurement?
- How is this test/measurement commonly described, if different from the technical term?
- Who is the target audience for this communication?

Data evaluability characteristics:

- What units are commonly used for reporting measurements to this audience?
- What is the direction of improvement (higher/larger always better, lower/smaller always better, optimal range)?
- What is the theoretical range of values?
- What range of observations is commonly observed?
- How much change in the value (in terms of the commonly used units) represents a clinically or practically significant change?
- Are there action or harm thresholds established or commonly used?
- Are there established target/goal levels or ranges?
    - If yes, is the target or goal the same or different for different people or observations?
- Is there an established categorization scheme for these values?
- Are historical values for this participant/source available?
- Are population norms available (e.g., mean, median, standard range, normal range)?
- Are there other key observations that this observation should be compared to (e.g., other locations)?
- Are percentile scores or other measures of relative position available?

Communication goal:

- Which of the following is most important for this audience to understand? (Pick ONE: result absolute level, result historical trend, result relationship to threshold, result categorization, result relative relationship to other observations, trends, or norms)

Most of the worksheet's questions ask the communicator to quantify an expert's knowledge in the form of relevant reference standards. For example, experts know over what range a given test result will commonly vary and what extreme outlier results are possible. In the medical domain, hemoglobin A1c test results (a measure of blood glucose control commonly assessed in patients with diabetes) and hematocrit results (a measure of oxygen carrying capacity in blood included in blood counts) are both percentages that could theoretically vary from 0% to 100%. However, in practice, hemoglobin A1c varies only between perhaps 4% and 12% (a value of 15% is a notable outlier), while hematocrit might vary from 18% to 42%. The worksheet therefore asks for the commonly observed range of values and a related question about the size of meaningful differences. In the above examples, a change of 0.5 percentage points in hemoglobin A1c is often considered a clinically significant difference, while a change of 0.5 percentage points in hematocrit is not particularly notable.

Other questions focus on the presence (or absence) of established reference standards. Are there action thresholds (i.e., levels that should trigger a recipient's attention) or goal ranges that are either formally or informally established? Examples include the U.S. Environmental Protection Agency's action threshold for household radon levels and its maximum contaminant levels (MCLs) for lead or mercury in drinking water. Similar guidelines in clinical practice govern the ordering of blood transfusions based on patients' blood hemoglobin levels (e.g., transfusion is recommended if a stable hospitalized patient's hemoglobin falls below 7 g/dL; Carson et al., 2016). Action thresholds and goal ranges are particularly valuable for communicating unfamiliar data because they explicitly communicate when something should be done (or does not need to be done) about a result. Categorization schemes (e.g., labeling results as "borderline high" or "very low") are also very useful, although it is not always clear how to translate a given category into action ("how low is too low?").

Note a tension exists between the range of variation implied by an action threshold and the range of variation observed when outliers are included. Take, for example, perfluoroalkyl (PFAS) and polyfluoroalkyl contaminants (PFOS). The current U.S. Environmental Protection Agency's health advisory level is 70 parts per trillion (ppt), and most water system observations fall far below those levels. Thus, the "commonly observed" range might be 0 to 100 ppt. Yet, affected communities have recorded levels that are 10 to 20 times higher than the advisory level. Scaling a visual display to include those outlier values would make all variations between 0 and 100 ppt seem small, yet it is precisely those variations that are often most important to communicate. For example, policy makers and residents need to distinguish in their minds between levels of 5 ppt and 50 ppt.

When action thresholds are not available, however, other reference standards can increase people's ability to make sense of their data. Population norms (i.e., average levels in the general population) can help people know whether they personally are comparatively better or worse than others. Historical data, especially personal norms (when available), are obviously particularly useful for clarifying changes over time. In the context of a particular research study, it can be helpful to clarify where a particular person or observation falls in the overall distribution. (This is analogous to the use of percentile scores to help students make sense of their college placement test results.) The problem with such relative comparison standards is that a result might be good compared with others in my community, yet nonetheless signals danger on an absolute basis (e.g., if everyone in a community has been exposed to a chemical at dangerous levels).

The last question on the worksheet, however, relates to communication goals, and it may be the most important one. The data characteristics worksheet is not a checklist but a list of types of context that may, or may not, be relevant in a given situation. Optimal communication design does not involve throwing every piece of contextual information into a message. Although providing reference ranges, categories, or thresholds can each improve the evaluability of unfamiliar data, providing all of the above undermines evaluability by creating confusion about which reference standard deserves attention (Scherer et al., 2018). Simplicity in visual display design ("less is more") can increase understanding (Peters, Dieckmann, Dixon, Hibbard, & Mertz, 2007; Tufte, 2001; Zikmund-Fisher, Fagerlin, & Ubel, 2010). Being aware of all of the different cues that experts use to derive meaning from a given number, however, is the first step to determining which type of comparison will most benefit a given audience.

One last note: Users of the worksheet should not limit themselves to answering only the questions that have a single, established answer. Some of the most useful reference information includes experts' judgments (e.g., of an action threshold) or the range of commonly observed values, which may be imprecise or variable across experts. If the goal is to help users to understand the data in the way that experts do, however, quantifying experts' knowledge and judgments adds to published standards.

Evaluability Versus Numeracy

Deficits in numeracy and graph literacy impede people's abilities to seek and use relevant quantitative health information (Galesic & Garcia-Retamero, 2011; Institute of Medicine, 2014; Lipkus & Peters, 2009; W. Nelson, Reyna, Fagerlin, Lipkus, & Peters, 2008). However, when communicators focus on how certain audience members may lack the relevant skills, the communicators can avoid considering whether their communications are optimally designed for everyone, even those with excellent numeracy skills, high levels of education, and access to resources. Talking about communication failures runs the risk of blaming the audience ("they didn't have the skills to understand it") rather than improving technique ("we didn't design the communication appropriately"). Communicators have a responsibility to design "for the way people are, not the way we wish they were" (Holly O. Witteman, personal communication, October 16, 2018).

In contrast to numeracy, information evaluability is a relevant concern regardless of audience. Framing the communicator's design task around information evaluability places attention on the characteristics of the data and the goal of the message. The communicator can then focus on creating the best message possible for the target audience, while acknowledging why the task may be difficult in a particular situation or for a particular audience or audience subgroup. The communicator can consider the following: Why are these numbers hard for *this* audience to evaluate? What type of evaluation is necessary for *this* user to accomplish *their* goals? Asking these questions brings the interests of the communicator and the audience into alignment, thereby increasing the chances of a communication that both parties would deem successful.

Information Evaluability Defines the Central Message of Data Communications

Innumerable individual and societal decisions depend on effective data communications (National Academies of Sciences, Engineering, and Medicine, 2017). Doctors and nurses share laboratory test results and risk statistics with patients with the hope of improving self-management and shared medical decision making. Environmental scientists provide measurements of global temperature changes to the public as part of discussions of climate change. Toxicologists relay the results of chemical exposure assessments to community residents and policy makers, who then seek to minimize their risks, determine appropriate policies, and so on. These decisions depend on patients, the public, and policy makers having an understanding of health-relevant data that incorporates lessons from experts' training and experience. Simply knowing the numbers is not enough. Decision makers must be able to use them.

The problem is that experts have a "Curse of Knowledge" problem (Heath & Heath, 2007). Experts, by definition, find the data that they work with on a daily basis to be easy to evaluate. In fact, the heart of their expertise is their prior understanding of relevant data characteristics and reference standards. The curse of experts' knowledge, however, is they often forget that their target audiences do not share their familiarity with data characteristics and context. This lack of awareness of audience knowledge underlies evaluability theory's third proposition that anyone who finds a particular type of data easy to evaluate will overestimate others' ability to interpret those data (Hsee & Zhang, 2010).

The solution is to recognize that health and science data communications, like all communications, should focus on an audience- and goal-specific purpose (D. E. Nelson, Hesse, & Croyle, 2009). While data sharing is about access to information, a communication has a specific audience and an intended message. Communicators are always considering (consciously or unconsciously) to whom they are speaking or writing, what they hope the audience will take away, and how to frame the message to achieve that goal. In data communication, identifying the central message means understanding why this audience *needs* these data and what *message* the audience needs to understand to accomplish their goals. In other words, what they need to think, feel, or do immediately upon receipt of this data communication. The last question in the Data Evaluability Characteristics Worksheet is designed to prompt identifying and prioritizing this central message.

Once data communicators know their audience and their goals, crafting an effective data communication is fundamentally an evaluability problem. The communicator needs to determine which specific contextual information will increase the evaluability of the data in ways that best align with the audience's needs. Each type of reference standard can sometimes be important, but the best choice depends on the situation. Ideally, the communicator finds the proverbial sweet spot by intentionally choosing the particular type of context that will be most of use to the target audience. The match to audience needs can be confirmed through interactions with representative audience members, such as cognitive interviews or usability tests.

Conclusion

Designing easily interpretable communications of scientific and health data is not easy. Few health professionals and scientists are trained in clear communications principles, and many times the people who most need to understand the data are those who are least familiar with their characteristics. Yet, framing the scientific data communication problem as being about information evaluability suggests simple ways to make many communications more effective.

Anyone communicating scientific or health measurements to nonexpert audiences (whether policy makers, patients, or the public more broadly) should include use-relevant contextual information to complement the measurement data themselves. It is never enough to simply provide access to data; a communicator must contextualize the measurements they provide to enable recipients to determine the data's good or bad implications the way an expert does. Doing so is not costly. It just requires careful attention to data characteristics, audience needs, and the way that experts construct their interpretations.

Increasing the evaluability of measurement data by intentionally providing use-relevant contextual information is the single best thing that a communicator can do to improve their communication effectiveness. This recommendation applies equally well to scientists wishing to inform policy makers, to government agencies seeking to inform the public, and to doctors communicating with patients. Doing so increases the chances that the patient, public, or policy maker audience does not just know what their numbers are but also what they mean.

Acknowledgments

The author thanks Holly Witteman, Jessica Ancker, Laura Scherer, and Christopher Hsee for their comments on an earlier version of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Brian J. Zikmund-Fisher https://orcid.org/0000-0002-1637-4176

References

American Hospital Association. (2016). *Individuals' ability to electronically access their hospital medical records, perform key tasks is growing*. Washington, DC: Author. Retrieved from http://www.webcitation.org/6t5QyAjYU

Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis*, *13*, 183-192. doi:10.1111/j.1539-6924.1993.tb01068.x

Carson, J. L., Guyatt, G., Heddle, N. M., Grossman, B. J., Cohn, C. S., Fung, M. K., . . . Tobian, A. A. R. (2016). Clinical practice guidelines from the AABB: Red blood cell transfusion thresholds and storage. *Journal of the American Medical Association*, *316*, 2025-2035. doi:10.1001/jama.2016.9185

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2007). "If I'm better than average, then I'm ok?" Comparative information influences beliefs about risk and benefits. *Patient Education and Counseling*, *69*, 140-144.

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

Gyrd-Hansen, D., Halvorsen, P., Nexøe, J., Nielsen, J., Støvring, H., & Kristiansen, I. (2011). Joint and separate evaluation of risk reduction: Impact on sensitivity to risk reduction magnitude in the context of 4 different risk information formats. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, *31*(1), E1-E10. doi:10.1177/0272989X10391268

Heath, C., & Heath, D. (2007). *Made to stick: Why some ideas survive and others die* (1st ed.). New York, NY: Random House.

Hibbard, J. H., & Peters, E. (2003). Supporting informed consumer health care decisions: Data presentation approaches that facilitate the use of information in choice. *Annual Review of Public Health*, *24*, 413-433. doi:10.1146/annurev.publhealth.24.100901.141005

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*, 247-257.

Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, *11*, 107-121.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*, 576-590. doi:10.1037/0033-2909.125.5.576

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*, 343-355. doi:10.1177/1745691610374586

Institute of Medicine. (2014). *Health literacy and numeracy: Workshop summary*. Washington, DC: The National Academies Press. doi:10.17226/18660

Krukowski, R. A., Harvey-Berino, J., Kolodinsky, J., Narsana, R. T., & DeSisto, T. P. (2006). Consumers may not use or understand calorie labeling in restaurants. *Journal of the American Dietetic Association*, *106*, 917-920. doi:10.1016/j.jada.2006.03.005

Lacey, H. P., Loewenstein, G., & Ubel, P. A. (2011). Compared to what? A joint evaluation method for assessing quality of life. *Quality of Life Research*, *20*, 1169-1177. doi:10.1007/s11136-011-9856-0

Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education & Behavior*, *36*, 1065-1081. doi:10.1177/1090198109341533

National Academies of Sciences, Engineering, and Medicine. (2017). *Communicating science effectively: A research agenda*. Washington, DC: The National Academies Press. doi:10.17226/23674

National Academies of Sciences, Engineering, and Medicine. (2018). *Returning individual research results to participants: Guidance for a new research paradigm*. Washington, DC: The National Academies Press. doi:10.17226/25094

National Research Council. (2010). *Advancing the science of climate change*. Washington, DC: The National Academies Press. doi:10.17226/12782

Nelson, D. E., Hesse, B. W., & Croyle, R. T. (2009). *Making data talk: Communicating public health data to the public, policy makers, and the press*. Oxford, UK: Oxford University Press.

Nelson, W., Reyna, V. F., Fagerlin, A., Lipkus, I. M., & Peters, E. (2008). Clinical implications of numeracy: Theory and practice. *Annals of Behavioral Medicine*, *35*, 261-274.

Newman, C. L., Howlett, E., & Burton, S. (2014). Shopper response to front-of-package nutrition labeling programs: Potential consumer and retail store benefits. *Journal of Retailing*, *90*, 13-26. doi:10.1016/j.jretai.2013.11.001

Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, *64*, 169-190.

Pinto-Prades, J. L., Robles-Zurita, J. A., Sánchez-Martínez, F.-I., Abellán-Perpiñán, J. M., & Martínez-Pérez, J. (2017). Improving scope sensitivity in contingent valuation: Joint and separate evaluation of health states. *Health Economics*, *26*(12), e304-e318. doi:10.1002/hec.3508

Ritov, I., Baron, J., & Hershey, J. C. (1993). Framing effects in the evaluation of multiple risk reduction. *Journal of Risk and Uncertainty*, *6*, 145-159. doi:10.1007/BF01065355

Scherer, A. M., Witteman, H. O., Solomon, J. B., Exe, N. L., Fagerlin, A., & Zikmund-Fisher, B. J. (2018). Improving understanding of test results by substituting (not adding) goal ranges: Experimental study. *Journal of Medical Internet Research*, *20*(10), e11027. doi:10.2196/11027

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics Press.

Zikmund-Fisher, B. J. (2013). The right tool is what they need, not what we have: A taxonomy of appropriate levels of precision in patient risk communication. *Medical Care Research and Review*, *70*(Suppl. 1), 37S-49S. doi:10.1177/1077558712458541

Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2004). "Is 28% good or bad?" Evaluability and preference reversals in health care decisions. *Medical Decision Making*, *24*, 142-148.

Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2010). A demonstration of "less can be more" in risk graphics. *Medical Decision Making*, *30*, 661-671.

Zikmund-Fisher, B. J., Scherer, A. M., Witteman, H. O., Solomon, J. B., Exe, N. L., & Fagerlin, A. (2018). Effect of harm anchors in visual displays of test results on patient perceptions of urgency about near-normal values: Experimental study. *Journal of Medical Internet Research*, *20*(3), e98. doi:10.2196/jmir.8889

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