Transparency in Information About Health –

Improving Medical Decision Making

D i s s e r t a t i o n
zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von

Dipl.-Psych. Nicolai Bodemer

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II

Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Gerd Gigerenzer
2. Prof. Dr. Matthias Ziegler
3. Prof. Dr. Michael Bosnjak

Tag der Einreichung: 20. Juni 2012

Tag der Verteidigung: 3. Dezember 2012

English Summary

This dissertation comprises four manuscripts focusing on health risk communication and medical decision making. The first manuscript discusses differences, commonalities, and the applicability of three major approaches to help patients make better decisions: nudging, social marketing, and empowerment. The second manuscript presents results of an evaluation of media coverage about the HPV vaccine of newspaper and Internet reports in Germany and Spain. Based on predefined standards for transparent, complete, and correct risk communication, the analysis revealed substantial shortcomings in how the media informed the public. The third manuscript centers on a standard format to communicate treatment benefits and harms: relative risk reductions and increases. Such formats have been found to misinform and mislead patients and health professionals. One suggestion is to always include information about baseline risk to reduce misunderstandings. Results show that even when baseline risk was communicated, it depended on the presentation format (percentage vs. frequency) and people's numeracy skills whether they correctly interpreted the risk reduction (or increase). Low numerates benefited from a frequency format, whereas high numerates performed better independent of the format. Yet, a substantial proportion of participants still misunderstood the meaning of a relative risk reduction (or increase). The fourth manuscript investigated how laypeople choose between medical treatments when ambiguity is present. One objection against communicating ambiguity is the claim that laypeople are ambiguity averse in the domain of gains and ambiguity seeking in the domain of losses. Results did not find supporting evidence for this claim in medical treatment choice. Moreover, most participants selected the same treatment option, independent of numeracy. However, the underlying choice strategies varied between individuals.

Keywords: Empowerment, Risk communication, Medical decision making, media analysis, Relative risk reduction, Ambiguity

Deutsche Zusammenfassung

Diese Dissertation umfasst vier Manuskripte zum Thema Risikokommunikation und medizinischen Entscheidungen. Das erste Manuskript diskutiert Unterschiede, Gemeinsamkeiten und die Anwendbarkeit von drei zentralen Ansätzen, die helfen sollen, bessere Entscheidungen zu treffen (Nudging, Social Marketing, Empowerment). Das zweite Manuskript präsentiert Ergebnisse einer Medienanalyse zur Evaluation von Zeitungs- und Internetberichten in Deutschland und Spanien über die HPV-Impfung. Basierend auf vordefinierten Standards für transparente, vollständige und korrekte Risikokommunikation, deckt die Medienanalyse Schwächen in der Berichterstattung auf. Das dritte Manuskript untersucht wie Laien relative Risikoreduktionen bzw. –erhöhungen, ein Standardformt in der Medizin, verstehen. Beide Formate führen Laien und Experten in die Irre und führen zur Überschätzung der tatsächlichen Effekte. Ein diskutierter Ausweg ist die zusätzliche Kommunikation der Basisrate. Die Ergebnisse zeigen, dass das Verständnis von relativen Risikoreduktionen (-erhöhungen) mit Basisrate von dem Präsentationsformat (Prozent- vs. Häufigkeitsformat) und der individuellen Fähigkeit im Zahlenverständnis abhängt. Teilnehmer mit geringem Zahlenverständnis profitierten von der Darstellung in Häufigkeiten; Teilnehmer mit hohem Zahlenverständnis zeigen ein besseres Verständnis unabhängig des Formats. Dennoch—selbst mit Basisrate—missverstehen viele Teilnehmer die Risikoinformation. Das vierte Manuskript untersucht wie Teilnehmer Behandlungen unter Unsicherheit auswählen. Ein Einwand gegen die Kommunikation von Unsicherheit ist die Behauptung, dass Menschen Unsicherheit in Gewinnsituationen vermeiden, in Verlustsituationen dagegen suchen. Die Ergebnisse dieser Studie in Bezug auf die Auswahl von medizinischen Behandlungen konnten diese Annahmen nicht bestätigen. Darüber hinaus wählte die Mehrheit der Teilnehmer die gleiche Behandlung, wenngleich sich die zugrundeliegende Auswahlstrategie unterschied.

Schlagwörter: Empowerment, Risikokommunikation, Medizinische Entscheidungen, Medienanalyse, Relative Risikoreduktion, Unsicherheit

Table of Contents

Chapter 1

General Introduction

Parts of this chapter are based on:
Bodemer, N., & Gaissmaier, W. (2012). Risk Communication in Health. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson (Eds.), *Handbook of Risk Theory* (S. 621–660). Dordrecht: Springer Netherlands.

General Introduction

Understanding health risks is a basic prerequisite for making health decisions. Policy-makers evaluate health risks to decide about the implementation of health programs, insurances assess the cost-effectiveness of health interventions, and doctors and patients need to know the statistics of benefits and harms of different treatment alternatives. In particular, the concepts of shared decision making and informed consent—that is, the mutual, interactive process between the doctor and the patient, who jointly make health decisions—challenge the classic paternalistic approach to the doctor-patient relationship. This requires the transparent communication of medical risks as a basis for informed decisions (Edwards & Elwyn, 2009; Gigerenzer & Gray, 2011; Gigerenzer, Gaissmaier, Kurz-Mielcke, Schwartz, & Woloshin, 2007). Others go even further and call for the "century of the patient" in health care to emphasize the importance of transparent information with far-reaching consequences for the individual as well as the health care system (Gigerenzer & Gray, 2011).

In my dissertation, I will present theoretical and empirical research on risk communication and medical decision making that gives insights into how we can help patients to understand health risks and make informed decisions. In this chapter, I will describe the framework of my dissertation and discuss its significance in the context of current research in the fields of medical risk communication and medical decision making.

Understanding Risks: Numeracy and intuitive design

The ability to deal with numbers in medical decision making is important for several reasons. It facilitates computation, encourages information search, improves interpretation of numerical information, facilitates the assessment of likelihood, and can increase or decrease involvement in medical decisions (Lipkus & Peters, 2009). The term numeracy has been introduced to describe "(…) *the degree to which individuals have the capacity to access, process, interpret, communicate, and act on numerical, quantitative, graphical, biostatistical, and probabilistic health information needed to make effective health decisions."* (Golbeck, Ahlers-Schmidt, Paschal, & Dismuke, 2005). It has been found that the public lacks

fundamental skills to deal with risks—a phenomenon that is not limited to laypeople, but has also been found in experts (Anderson, Gigerenzer, Parker, & Schulkin, 2012; Reyna, Nelson, Han, & Dieckman, 2009). Does this undermine the ideals of shared decision making and informed consent? The answer is no; there are at least two complementary solutions to overcome *innumeracy*. The first solution is training. School curricula should include programs that would teach children to understand statistical information already in early grades. Moreover, health professionals should receive training about different methods of risk communication and the best ways of presenting health statistics to their patients. The second solution is intuitive design of statistical information. Presenting risks intuitively, that is using the principles of ecological rationality, can help overcome people's difficulties in understanding of statistics (Todd, Gigerenzer, & the ABC Research Group, 2012). For example, people often appear to have difficulties in solving Bayesian inference tasks, such as computing the probability of a woman having breast cancer given a positive mammogram, based on the hit rate, sensitivity and false alarm rate of the screening test. However, this difficulty was mainly observed when the relevant information was presented in conditional probabilities. Presenting the same information in natural frequencies facilitated Bayesian inferences—a format that simplifies the calculation for the human mind (Gigerenzer & Hoffrage, 1995). Hence, when risks are presented intuitively, that means, when the external format matches the structure of the human mind's core capacities, we can improve understanding.

What is "good" risk communication?

The major challenge in health risk communication is discovering how to help patients in making better decisions. But what is the objective of risk communication? What is the standard by which risk communication should be evaluated? Some argue that the ultimate goal of risk communication is to change people's (health) behavior towards what is considered "optimal" (Andreasen, 1995; Lee & Kottler, 2011; Thaler & Sunstein, 2009). For instance, a health campaign might aim at increasing participation rates in cancer screening programs such as mammography. Others propose that, because such an "optimal" behavior is

difficult to define, the knowledge and education of the patient should be the major outcome (Gigerenzer & Gray, 2011). This means that one should primarily aim at equipping patients with knowledge. One example is informing women about the statistical evidence of mammography screening so they themselves can decide whether to participate or not.

In Chapter 2, I will present a theoretical paper in which we compare three common approaches proposed to improve health decisions —nudging, social marketing and empowerment—and discuss their applicability in health care. These approaches differ in terms of their underlying assumptions about people's decision making, their aims and the strategies implemented to achieve them. Whereas it is often assumed that these approaches are mutually exclusive (e.g., Ariely, 2010; Marteau, Ogilvie, Roland, Suhrcke, & Kelly, 2010; Thaler & Sunstein, 2009), I argue that they are different tools of the same toolbox. None of the tools is always best. Rather, its applicability depends on the problem. I suggest that the major challenge is to find out when to use which tool.

The current practice in health care

Involving patients in the medical decision process requires transparent communication of health statistics. Does the current practice in risk communication conform to this ideal? Gigerenzer and Gray (2011) identified "the seven sins in healthcare", the interplay of seven factors that produce misinformed and misled patients, resulting in poor health decisions: biased funding of research, biased reporting in medical journals, biased reporting in health pamphlets, biased reporting in the media, commercial conflicts of interest, defensive decision making, and doctors' lack of understanding of health statistics. Three of these seven sins emphasize the problem of biased reporting within the medical community as well as in direct-to-consumer communication. Biased reporting refers to two phenomena. First, evidence is presented incompletely and one-sided. For instance, while the benefit of mammography screening on breast cancer mortality reduction is usually communicated, the potential harms of the treatment, such as overdiagnosis and overtreatment, are mostly omitted (Gigerenzer et al., 2007). Second, evidence is often framed in non-transparent and incomprehensible formats

that mislead patients to over- or underestimate the actual underlying effects. Consequences of biased reporting are manifold: patients dramatically overestimate the benefit of cancer screening programs (Gigerenzer, Mata, & Frank, 2009), patients and health professionals ignore the possibility of false alarms, resulting in overdiagnosis, overtreatment, and expensive follow-up care, anxieties and fears (Gigerenzer et al., 2007; Lafata et al., 2004).

In Chapter 3, I will present a study in which I extended previous views on media reporting. In a cross-cultural comparison between Germany and Spain, two countries with different health systems, I compared newspaper and Internet reports about the human papillomavirus (HPV) vaccination. Based on predefined standards for transparent risk communication, I evaluated media reports with respect to transparency, completeness and correctness of information. Monitoring the current practice of risk communication is important for two reasons. First, the media are the most prominent channel to communicate medical innovations and treatments to the public (Grilli, Ramsay, & Minozzi, 2009). Second, to understand patient's decision making, it is essential to consider the ecological setting in which decisions take place (Todd et al., 2012). The study revealed shortcomings in the current practice of risk communication in both countries and proposes ways to improve media reporting.

"Debiasing" nontransparent risk communication: Relative risk reductions with baseline risk

Biased reporting is also common in medical journals. One prominent example is the use of relative risk reductions to present treatment benefits. For instance, out of 222 abstracts in leading medical journals in 2003-2004, 68% failed to state the absolute risk (Sedrakyan & Shih, 2007). In 2009, 16 out of 37 articles in the British Medical Journal did not report the underlying absolute risks (Gigerenzer, Wegwarth, & Feufel, 2010). This is problematic because laypeople and health professionals overestimate treatment effects when presented in relative risk reductions (Akl et al., 2012; Covey, 2007; Cranney & Walley, 1996; Edwards, Elwyn, Covey, Matthews, & Pill, 2001). One suggestion to resolve this bias is to consistently communicate relative risk reductions with baseline risk (Natter & Berry, 2005).

In Chapter 4, I will present a paper that shows that the interpretation of relative risk reductions (and increases) with baseline risks depends on two factors: the presentation format of the baseline risk (percentage vs. frequency) and people's numeracy skills. Presenting baseline risk in frequency format helped in particular low numerates to correctly interpret relative risk reductions. High numerates performed better independently of the format. However, even when baseline risk was included, a large proportion misinterpreted relative risk reductions leading to overestimation of treatment effects. Hence, we propose an alternative and more transparent way to communicate benefits and harms.

Transparent risk communication: The role of ambiguity

Gigerenzer and colleagues (2007) coined the concept of *statistical literacy* and proposed a minimum set of requirements needed to adequately deal with medical information (for an overview, see Chapter 3). One central component of statistical literacy is the ability to deal with uncertainty. In contrast to risks, which are generally measureable, uncertainties are not (Knight, 1921). However, a less strict definition of uncertainty allows estimating parameters of uncertainty, like ranges or confidence intervals (Politi, Col, & Han, 2007). The term *ambiguity* has been introduced to refer to this sub concept of uncertainty (Han, Klein, Lehman, Massett, Lee, & Freedman, 2011). Ambiguity plays a major role in medical decision making, because evidence is (i) usually limited, due to constraints in the research design (e.g., sample characteristic, reliability and validity of measures), (ii) based on population data, and hence difficult to be applied to individuals and (iii) based on past events and assume that factors in the future do not change (Politi et al., 2007). Ambiguity is rarely, if ever, included in medical risk communication. One reason is that it is assumed that people have difficulties in dealing with ambiguity and react to it with aversion (Epstein, 1999; Frewer, Hunt, Brennan, Kuznesof, Ness, & Ritson, 2003).

In Chapter 5, I will present a study in which participants had to choose between a certain and an ambiguous treatment option. Participants were presented treatment effects about benefits or harms as either a point estimate (the average rate of benefits or harms, such

as 20 out of 100 experience the effect) or a range (the lower and upper bound, such as between 10 – 30 out of 100 experience the effect). I investigated how ambiguity influences treatment choice and which information participants use when being confronted with a range (i.e., the size of the range, its upper and lower bounds, or its midpoint). In general, I did not find evidence for ambiguity aversion, and observed high heterogeneity in people's choice strategies. Moreover, I found that participants' preference of a treatment option was rather independent of its degree of ambiguity. However, small differences were found depending on whether treatment benefits or harms were presented, and on people's numeracy skills. Results suggest that ambiguity is an important element for a decision maker and that its communication increases completeness and transparency in risk communication.

In sum, in four manuscripts I apply psychological theories and methods to help patients make better medical decisions. These range from general theories about rationality (Chapter 2), their application to standards for transparent risk communication based on principles of ecological rationality and intuitive design (Chapter 3), findings and research designs from cognitive psychology to understand patient reasoning (Chapter 4), and challenging classical concepts on decision making under ambiguity (Chapter 5). Thereby, I extend the current knowledge and propose future research directions based on this dissertation's findings (Chapter 6).

References

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., Costiniuk, C., et al. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews (Online)*, (3), CD006776. doi:10.1002/14651858.CD006776.pub2

Anderson, B. L., Gigerenzer, G., Parker, S., & Schulkin, J. (2012). Statistical literacy in obstetricians and gynecologists. *Journal for Healthcare Quality* doi:10.1111/j.1945-1474.2011.00194.x

Andreasen, A. R. (1995). *Marketing social change: Changing behavior to promote health, social development, and the environment*. San Francisco: Jossey-Bass.

Ariely, D. (2010). *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins.

Covey, J. (2007). A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making*, *27*(5), 638–654. doi:10.1177/0272989X07306783

Cranney, M., & Walley, T. (1996). Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *The British Journal of General Practice*, *46*(412), 661–663.

Edwards, A, Elwyn, G., Covey, J., Matthews, E., & Pill, R. (2001). Presenting risk information—a review of the effects of „framing“ and other manipulations on patient outcomes. *Journal of Health Communication*, *6*(1), 61–82. doi:10.1080/10810730150501413

Edwards, Adrian, & Elwyn, G. (2009). *Shared decision-making in health care: Achieving evidence-based patient choice*. Oxford; New York: Oxford University Press.

Epstein, L. G. (1999). A definition of uncertainty aversion. *Review of Economic Studies*, *66*(3), 579–608. doi:10.1111/1467-937X.00099

Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research*, *6*(1), 75–85. doi:10.1080/1366987032000047815

Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of the National Cancer Institute*, *101*(17), 1216–1220. doi:10.1093/jnci/djp237

Gigerenzer, G., Wegwarth, O., & Feufel, M. (2010). Misleading communication of risk. *British Medical Journal*, *341*, c4830. doi:10.1136/bmj.c4830

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. doi:10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Gray, M. (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge: MIT Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Golbeck, A. L., Ahlers-Schmidt, C. R., Paschal, A. M., & Dismuke, S. E. (2005). A definition and operational framework for health numeracy. *American Journal of Preventive Medicine*, *29*(4), 375–376. doi:10.1016/j.amepre.2005.06.012

Grilli, R., Ramsay, C., & Minozzi, S. (2002). Mass media interventions: effects on health services utilisation. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. doi.wiley.com/10.1002/14651858.CD000389

Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2011). Varieties of uncertainty in health care: A conceptual taxonomy. *Medical Decision Making*. doi:10.1177/0272989X10393976

Knight, F.H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Houghton Mifflin Co.

Kotler, P., & Lee, N. (2012). *Social marketing: Influencing behaviors for good*. Los Angeles: Sage Publications.

Lafata, J. E., Simpkins, J., Lamerato, L., Poisson, L., Divine, G., & Johnson, C. C. (2004). The economic impact of false-positive cancer screens. *Cancer Epidemiology, Biomarkers & Prevention*, *13*(12), 2126–2132.

Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education & Behavior*, *36*(6), 1065–1081. doi:10.1177/1090198109341533

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: Can nudging improve population health? *British Medical Journal*, *342*, d228. doi:10.1136/bmj.d228

Natter, H. M., & Berry, D. C. (2005). Effects of presenting the baseline risk when communicating absolute and relative risk reductions. *Psychology, Health & Medicine*, *10*(4), 326–334. doi:10.1080/13548500500093407

Politi, M. C., Han, P. K. J., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695. doi:10.1177/0272989X07307270

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. doi:10.1037/a0017327

Sedrakyan, A., & Shih, C. (2007). Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. *Medical Care*, *45*, 23–28. doi:10.1097/MLR.0b013e3180642f69

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. Oxford; New York: Oxford University Press.

Chapter 2

Finding the right tool to improve health decisions: Nudging, social marketing, or empowerment

Markus Feufel[1,2] & Nicolai Bodemer[1,2]

[1]Harding Center for Risk Literacy, Berlin, Germany
[2]Max Planck Institute for Human Development, Berlin, Germany

[Word count: 1836]

Feufel, M.A. & Bodemer, N. (2012). *Finding the right tool to improve health decisions: Nudging, social marketing, empowerment*. Manuscript in preparation.

Abstract

One way to improve resource allocation in health care is to help patients and health professionals make better decisions. Two main approaches have been put forth: nudging and social marketing. Our analysis shows that these strategies apply only when there is (1) good evidence for a particular treatment and no legitimate alternatives exist or (2) a strong social agreement as to what is considered a desirable decision. If these conditions are not met, we argue that a third strategy needs to be added to the toolbox—empowerment based on education and transparent communication—to allow health professionals and patients compare risks and benefits across treatments and assume their responsibilities as partners in negotiating care decisions. To better harness the power of the available public health strategies, we specify when to apply which of the three strategies to facilitate better health decisions.

Introduction

Modern health systems are plagued with the tenacious problem of unwarranted practice variation, that is, underuse, overuse, and misuse of resources (Goodlee, 2011; Wennberg, 1984). How can this challenge be tackled? One avenue to improving resource allocation is to help patients and health professionals make better decisions and ultimately do what is "best" for them and society (Andreasen, 1995; Gigerenzer & Gray, 2011; Thaler & Sunstein, 2009). The two most common approaches are nudging (Marteau, Ogilvie, Roland, Suhrcke, & Kelly, 2011; Thaler & Sunstein, 2009) and social marketing (Andreasen, 1995; Evans, 2006). Based on an analysis of their underlying assumptions, we argue that these approaches should be applied only if one of the following conditions is met: (a) there is good evidence for a dominant therapeutic or diagnostic procedure and no alternatives are available or (b) there is a "strong social agreement" as to what is considered a good decision such as agreeing to be a potential organ donor or staying at home when sick with a contagious disease. If neither of these conditions is met, a third strategy needs to be added to the public health toolbox—empowerment of patients and health professionals based on education and transparent communication—to enable the negotiation of desirable decisions and care solutions (Gigerenzer & Gray, 2011). We first outline the nudging and social marketing approaches. Then we discuss their assumptions to elaborate why a third strategy is needed and introduce particular empowerment strategies. Finally, we specify which strategies to apply for what kind of care decision.

Comparison of the three major approaches

Nudging

The method used to nudge people—designing environments that afford certain actions—has a long tradition in the field of human factors and ergonomics (Norman, 1988). Its rationale differs from traditional approaches to environmental design, however. Nudging assumes that people's choices must be guided because mental capacity limitations result in decisions that are systematically inferior to the normative standards implied by those who nudge. Rather than elaborating rationales, raising awareness, or educating people, the quality

of health decisions is therefore improved by designing environments that facilitate good and deter inferior decisions (Ariely, 2010; Thaler & Sunstein, 2009). An example is to change opt-in to opt-out defaults to increase registration rates for organ donation as a socially desirable behavior (Johnson & Goldstein, 2003).

Social marketing

Social marketing is based on the assumption that "if marketing can encourage us to buy a Ferrari, it can persuade us to drive safely" (Hastings, 2006). Its goal is "to influence health behavior" and raise awareness about (socially) desirable behaviors in specific populations, by targeting their values and motivations via appropriate communication channels (Evans, 2006). Channels may range from the mass media to message placements, community outreach, and interpersonal communication. Examples include social marketing strategies to contain the spread of flu germs across the population as in the NHS's "Catch it, bin it, kill it" campaign or to increase registration rates for organ donations.

Empowerment

Both nudging and social marketing are based on the assumption that there are well-defined objective or social criteria for good health decisions that are agreed upon by both the public and the designers of nudging and social marketing campaigns. Assumed and patient relevant criteria may not always coincide, even if they appear to be objective and clear-cut. Consider the example of an American Indian patient with hip dysplasia (Rabin, Barnett, Arnold, Freiberger, & Brooks, 1965). To avoid walking with a limp and risk long-term osteoarthritis due to increased wear, an initiative similar to "Doctors without Borders" surgically resolved the patient's hip dysplasia. Although the diagnosis and procedures seemed medically correct to both physicians and the patient, the operation ultimately failed relative to the patient's cultural values and norms. A common side effect of the surgery—reduced hip function—thwarted his ability to meet with his tribe (a cultural value) sitting cross-legged in a circle on the ground (a cultural norm).

The key issue is that nudging and social marketing are primarily concerned about behavioral change—thereby both approaches are often silent about the underlying rationale for such a change. This hinders patients to question the direction into which a behavior is supposed to be guided. Why is that? The reason therefore lies in a more implicit second assumption of these approaches: suboptimal decisions relate to a lack of mental abilities such as cognitive biases (nudging) or a lack of knowledge and motivation (social marketing). In fact, many patients know surprisingly little about basic medical facts and procedures (Bachmann, Gutzwiller, Puhan, Steurer, Steurer-Stey, & Gigerenzer, 2007) and lack basic skills for accessing high quality medical information (Feufel & Stahl, 2012). Also, both patients and health professionals have problems understanding health statistics because they are provided with non-transparent information sources (Bodemer, Müller, Okan, Garcia-Retamero, & Neumeyer-Gromen, 2012). However, there is growing evidence that this can be changed. Research has shown that transparent communication can improve comprehension and knowledge of health statistics (Bodemer & Gaissmaier, 2012; Gigerenzer, Gaissmaier, Kurz-Mielcke, Schwartz, & Woloshin, 2007), self-care behaviors (Feufel, Schneider, & Berkel, 2010), and even increase adherence to public health campaigns (Schneider, Feufel, & Berkel, 2011). Thus, whether the assumption of lacking abilities, knowledge, or motivation holds depends on how medical evidence and how information is communicated.

So what if the assumptions underlying nudging and social marketing are not met? If there are multiple normative criteria as in the example of an American Indian with hip dysplasia, patients and health professionals must negotiate their decisions. This requires that people have the skills to find and evaluate medical evidence and are provided with transparently formatted information. To *empower* patients and health professionals to negotiate good health decisions, a two-pronged approach is necessary (Feufel, Antes, Nelson, Gigerenzer, Gray, & Mäkela, 2011): First, environmental (i.e., information) design to inform patients transparently about self-care behaviors (Feufel et al., 2010; Schneider et al., 2011) and both patients and health care professionals about the benefits and harms of available

diagnostic and treatment options (Bodemer & Gaissmaier, 2012; Gigerenzer et al., 2007); Second, educational interventions to provide patients and health professionals with the basic skills to find and evaluate quality medical evidence (Feufel & Stahl, 2012; Gigerenzer et al., 2007). Hence, in contrast to nudging and social marketing, patient knowledge is the primary objective—behavioral change may or may not be a consequence.

When to apply which strategy?

A toolbox of strategies is useful only if one knows when to apply which tool. In this section, we review the applicability of nudging, social marketing, and empowerment with respect to three major types of care that have been identified in the literature (Wennberg, 1984): efficient care, preference-sensitive care, and supply-sensitive care (for a summary see Table 1).

Efficient care

No doubt, there are situations where efficient care practices should be applied to achieve optimal care outcomes and avoid medical errors. For instance, to avoid severe infections from central vein catheters, simple checklists of hygiene measures have been proven effective and should therefore be followed (Pronovost et al., 2010). Nudging and social marketing approaches seem to be well suited to implement efficient care practices in medical settings. However, many medical decisions have to be made under considerable uncertainty, not simply related to limited or conflicting evidence but to the supply of resources and expertise as well as patient preferences.

Preference-sensitive care

If legitimate treatment alternatives exist, involving tradeoffs among possible treatments and outcomes, care is preference-sensitive and the “best” decision cannot be made without considering patient preferences (O’Connor, Llewellyn-Thomas, & Flood, 2004). In such cases, prerequisites for nudging and social marketing are not met and it is ethically questionable to nudge or apply social marketing without informing about the available

options, ways to implement them as well as their benefits and harms. For preference-sensitive care, transparent communication and understanding of the available and lacking evidence is the necessary basis for informed health decisions (Gigerenzer & Gray, 2011).

Supply-sensitive care

Since the early 1980s, Jack Wennberg and colleagues have identified unwarranted practice variations in the U.S. such as tonsillectomy rates varying between 8% and 70% among children in the state of Vermont. According to Wennberg, these variations reflect Roemer's law: greater supply of resources and experts tends to result in more physician visits, testing, and in-patient treatments, independent of medical need (Dartmouth Atlas Project, 2007). Moreover, patients in regions with higher utilization rates due to greater supply show slightly increased mortality and lower satisfaction with care (Fisher, Wennberg, Stukel, Gottlieb, Lucas, & Pinder, 2003). If more care is not better, there should be strong social agreement to improve resource allocation. Thus, nudging and social marketing campaigns can and should be applied to stop health professionals and private corporations maximize supply usage rates and profits. Empowerment approaches must further support these efforts by making supply-induced variations transparent and accessible to health professionals and patients so they can compare and adapt care delivery practices (Mulley & Wennberg, 2011).

A matter of target

Independent of which kind of care is at issue, the choice of the tool depends on which objective a communicator pursues. Shared decision making and informed consent focus on knowledge as the major outcome, but a communicator might be primarily concerned about changing behavior, independently of whether the patient is informed or not.

Table 1

Assumptions, Tools, and Applicability of Three Major Approaches to Improving Health Decisions.

	Assumptions	Tools	Applicability
Nudging	1. There is a normative decision criterion, either supported by evidence or social agreement 2. Lack of mental capacities warrant behavioral guidance	1. Environmental design to facilitate or deter (un)desirable behaviors without providing rationales	1. Efficient care 2. Supply-sensitive care
Social Marketing	1. There is a normative decision criterion, either supported by evidence or social agreement 2. Lack of knowledge or motivation warrant behavioral guidance	1. Knowledge/awareness campaigns provide rationales to advertise or discourage (un)desirable behaviors	1. Efficient care 2. Supply-sensitive care
Empowerment	1. There are multiple decision criteria so patients and health professionals must negotiate medical decisions	1. Environmental design to facilitate comprehension 2. Knowledge/awareness campaigns to support negotiation of decisions	1. Efficient care 2. Preference-sensitive care 3. Supply-sensitive care

Conclusions

The three approaches to better medical decisions we described in this article – nudging, social marketing, and empowerment – are not mutually exclusive, but part of one toolbox aimed at improving health decisions (see Table 1). Just like you may choose between front, rear, or four-wheel drive to maneuver unwieldy terrain, interventions aimed at improving health system functioning should change both the health care system and support its patients and health professionals (Feufel et al., 2011). Whereas nudging implies health system redesign to subtly prompt desirable and deter undesirable behaviors, social marketing aims at changing individuals' choices directly by campaigning for and against certain health behaviors. Neither one of these approaches will suffice by itself. We have argued that nudging and social marketing approaches are suitable if there is (a) good evidence for a particular

treatment and no legitimate alternatives exist or (b) a strong social agreement as to what is considered a desirable decision and behavioral change is to be achieved rather independent of patient involvement. In most (other) cases, patients and physicians must negotiate their decisions. One necessary requirement is therefore to empower patients and health professionals; first by educating patients how they can find and evaluate quality medical evidence, and second, by providing transparent information about the available (or missing) evidence, self-care behaviors, and supply-induced usage patterns.

Nudging and social marketing are popular tools to facilitate better decisions in patients and health professionals. The analysis of their underlying assumptions and methods showed that we must enrich the available toolbox by introducing a third alternative, empowerment, that works when the others fail and guide tool selection based on the kind of care decisions patients and health professionals face. With the right tool in hand we will be better able to harness their powers and ultimately improve health decisions.

References

Andreasen, A. R. (1995). *Marketing social change: Changing behavior to promote health, social development, and the environment*. San Francisco: Jossey-Bass.

Ariely, D. (2010). *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins.

Bachmann, L. M., Gutzwiller, F. S., Puhan, M. A., Steurer, J., Steurer-Stey, C., & Gigerenzer, G. (2007). Do citizens have minimum medical knowledge? A survey. *BMC Medicine*, *5*(1), 14. doi:10.1186/1741-7015-5-14

Bodemer, N., & Gaissmaier, W. (2012). Risk Communication in Health. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson (Eds.), *Handbook of Risk Theory* (S. 621–660). Dordrecht: Springer Netherlands.

Bodemer, N., Müller, S. M., Okan, Y., Garcia-Retamero, R., & Neumeyer-Gromen, A. (2012). Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine. *Vaccine*, *30*(25), 3747–3756. doi:10.1016/j.vaccine.2012.03.005

Dartmouth Atlas Project Topic Brief (2007). *Supply-sensitive care*. [cited 2012 March 15]. Retrieved from: http://www.dartmouthatlas.org/downloads/reports/supply_sensitive.pdf

Feufel, M. A., Schneider, T. R., & Berkel, H. J. (2010). A field test of the effects of instruction design on colorectal cancer self-screening accuracy. *Health Education Research*, *25*(5), 709–723. doi:10.1093/her/cyq015

Feufel, M.A., Antes, G., Nelson, D., Gigerenzer, G., Gray, J.A.M., Mäkela, M., et al. (2011). How to achieve better health care: Better systems, better patients, or both? In: G. Gigerenzer G & J.A.M. Gray (Eds.). *Better doctors, better patients, better decisions: Envisioning healthcare 2020* (pp. 117-134.). Cambridge: MIT Press.

Feufel, M. A . & Stahl, S. F. (2012). What do web-use skill differences imply for online health searches?. *Journal of Medical Internet Research, 14*, e87

Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. J., Lucas, F. L., & Pinder, E. L. (2003). The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine*, *138*(4), 288–298.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. doi:10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Gray, J.A.M.. (2011). Launching the century of the patient. *Better doctors, better patients, better decisions: Envisioning health care 2020* (pp. 3–28). Cambridge, MA: MIT Press.

Godlee, F. (2011). Tackling practice variation. *British Medical Journal*, *342*, d1884. doi:10.1136/bmj.d1884

Hastings, G. (2006). Putting social marketing into practice. *British Medical Journal*, *332*(7551), 1210–1212. doi:10.1136/bmj.332.7551.1210

Johnson, E. J. & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*(5649), 1338–1339. doi:10.1126/science.1091721

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: Can nudging improve population health? *British Medical Journal*, *342*, d228. doi:10.1136/bmj.d228

Mulley, A. G. & Wennberg, J. E. (2011) Reducing Unwarranted Variation in Clinical Practice by Supporting Clinicians and Patients in Decision Making. In: G. Gigerenzer & J.A.M. Gray JAM, editors. *Better doctors, better patients, better decisions: Envisioning healthcare 2020* (pp. 45-52). Cambridge: MIT Press.

O'Connor, A. M., Llewellyn-Thomas, H. A., & Flood, A. B. (2004). Modifying unwarranted variations in health care: shared decision making using patient decision aids. *Health affairs (Project Hope)*, *Suppl Variation*, VAR63–72. doi:10.1377/hlthaff.var.63

Pronovost, P. J., Goeschel, C. A., Colantuoni, E., Watson, S., Lubomski, L. H., Berenholtz, S. M., Thompson, D. A., et al. (2010). Sustaining reductions in catheter related bloodstream

infections in Michigan intensive care units: observational study. *British Medical Journal*, *340*, c309. doi:10.1136/bmj.c309

Rabin, D.L., Barnett, C.R., Arnold, W.D., Freiberger, R.H., Brooks, G. (1965). Undertreated congenital hip disease. *American Journal of Public Health and the Nations Health, 55*,3–44.

Schneider, T. R., Feufel, M. A., & Berkel, H. J. (2011). Promoting Colorectal Cancer Screening in Public Health Outreach Campaigns. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 637–646. doi:10.1177/0018720811427134

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

Wennberg, J. E. (1984). Dealing with medical practice variations: A proposal for action. *Health Affairs*, *3*(2), 6–32. doi:10.1377/hlthaff.3.2.6

Chapter 3

Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine

Nicolai Bodemer[1, 2], Stephanie M. Müller[3], Yasmina Okan[4], Rocio Garcia-Retamero[4], & Angela Neumeyer-Gromen[1, 2]

[1]Harding Center for Risk Literacy, Berlin, Germany
[2]Max Planck Institute for Human Development, Berlin, Germany

[3]Center for Empirical Research in Economics and Behavioral Science (CEREB), University of Erfurt, Erfurt, Germany

[4] University of Granada, Granada, Spain

[Word count: 3608]

Bodemer, N., Müller, S. M., Okan, Y., Garcia-Retamero, R., & Neumeyer-Gromen, A. (2012). Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine. *Vaccine*, *30*, 3747–3756.

Abstract

The media is a powerful tool for informing the public about health treatments. In particular, the Internet has gained importance as a widely valued source for health information for parents and adolescents. Nonetheless, traditional sources, such as newspapers, continue to report on health innovations. But do websites and newspaper reports provide balanced information? We performed a systematic media analysis to evaluate and compare media coverage of the human papillomavirus (HPV) vaccine on websites and in newspapers in Germany and Spain. We assessed to what extent the media provide complete (pros and cons), transparent (absolute instead of relative numbers), and correct information about the epidemiology and etiology of cervical cancer as well as the effectiveness and costs of the HPV vaccine. As a basis for comparison, a fact box containing current scientific evidence about cervical cancer and the HPV vaccine was developed. The media analysis included 61 websites and 141 newspaper articles in Germany, and 41 websites and 293 newspaper articles in Spain. Results show that 57% of German websites and 43% of German newspaper reports communicated correct estimates of epidemiological data, whereas in Spain 39% of the websites and 20% of the newspaper did so. While two thirds of Spanish websites explicitly mentioned causes of cervical cancer as well as spontaneous recovery, German websites communicated etiological information less frequently. Findings reveal that correct estimates about the vaccine's effectiveness were mentioned in 10% of German websites and 6% of German newspaper reports; none of the Spanish newspaper reports and 2% of Spanish websites reported effectiveness correctly. Only German websites (13%) explicitly referred to scientific uncertainty regarding the vaccine's evaluation. We conclude that the media lack balanced reporting on the dimensions completeness, transparency, and correctness. We propose standards for more balanced reporting on websites and in newspapers.

Introduction

The media play an important role in informing and educating the public about health interventions (Chapman & Lutton, 1994; James, James, Davies, Harvey, & Tweedle, 1999). In particular, the Internet seems to replace traditional sources such as newspapers and pamphlets (Hesse, Nelson, Kreps, Coyle, Arora, Rimer, et al., 2005). For instance, up to 47% of European Internet users look for health-related information online (European Commission, 2011). About two thirds of U.S. Internet users seek online information about diseases and symptoms, and about half look for information on prescription drugs (Fox, 2011). Parents preferred the Internet as the second most important source after health care providers for health-related information for their children (D'Alessandro, Kreiter, Kinzer, & Peterson, 2004; Khoo, Bolt, Babl, Jury, & Goldman, 2008). Adolescents also obtain health information on the Internet (Gray, Klein, Noyce, Sesselberg, & Cantrill, 2005): they show high familiarity, competency and comfort with this medium, and the Internet assures anonymity when queries are about sexuality and related risks.

Despite the increasing impact of the Internet, newspapers remain an important channel for disseminating information about public health issues (Slater, Long, Bettinghaus, & Reineke, 2008; Stryker, Moriarty, & Jensen, 2008): newspapers reach a large segment of the population, across ages and without regard to Internet accessibility. They offer a trustworthy and usually concise alternative to the Internet, without requiring active information search. Especially people lacking basic media literacy skills have difficulty identifying reliable and trustworthy information and feel easily overwhelmed by the Internet (Morahan-Martin, 2004). Both media types represent an important public resource for health news, treatments, and innovations, but they differ on accessibility, target population, and information presentation**.**

Media analysis of the HPV vaccine

We conducted a content analysis of the media coverage of the human papillomavirus (HPV) vaccine. We aimed to evaluate websites and newspaper reports about the vaccine in two countries—Germany and Spain. Several reasons make the HPV vaccine a suitable subject

for a media analysis: First, it has received extensive media coverage in several countries due to its innovative application for cancer prevention and its public health relevancy (targeting girls 12–17 years old), and critical voices have questioned the vaccine's introduction (Dören et al., 2008; Martin-Llaguno & Alvarez-Dardet, 2010). Second, information about the vaccine addresses parents and adolescents; both groups seek health information on the Internet and in newspaper reports.

Previous media analyses of the HPV vaccine have already pointed out a lack of basic information about risk factors, transmission, and symptoms (Abdelmutti & Hofman-Goetz, 2009; Habel, Liddon, & Stryker, 2009; Kelly, Leader, Mittermaier, Hornik, & Cappella, 2009). Tozzi and colleagues (2010) found higher accessibility, credibility, and content ratings on English websites compared with Italian websites. Additionally, public health and university websites scored higher on credibility, content, and design compared with private or company websites. We extended previous research by comparing reports from two different media types, newspapers and Internet websites, in two countries whose media coverage about the HPV vaccine has not yet been assessed: Germany and Spain. We assessed websites and newspapers because (1) their target groups differ in age, income, and education (Cotton & Gupta, 2004), (2) the Internet represents a rather new way to inform about health, and (3) Internet sources often provide information tailored for specific groups, whereas newspaper reports address a broader audience (Schönbach, de Waal, & Lauf, 2005). Previous analyses assessed the quality and readability of websites (Tozzi et al., 2010) and the completeness of information in different media types (Abdelmutti & Hoffman-Goetz, 2009; Habel et al., 2009; Kelly et al., 2009). Our analysis focused on the evaluation of the content: we assessed what information was provided, whether the information was based on evidence, and what format was used to communicate this information. Our main objective was to assess how basic information about epidemiology, etiology, and benefits and harms of the HPV vaccine is communicated. In particular, we assessed the prevalence of biased reporting about the HPV vaccine.

The risks of biased reporting

Biased reporting is a real risk in the media's coverage of health issues (Gigerenzer & Gray, 2011; Gigerenzer, Wegwarth, & Feufel, 2010). Bias can result from providing incomplete information (e.g., omitting drawbacks) or using nontransparent information formats (e.g., reporting relative instead of absolute risk reductions). Several media analyses have documented shortcomings of media health coverage (Frost, Frank, & Maibach, 1997; Kurzenhäuser, 2003; Moynihan et al., 2000; Steckelberg, Balgenorth, & Mühlhauser, 2001). Often, important numerical information is lacking or fragmented (Gigerenzer, Gaissmaier, Kurz-Mielcke, Schwartz, & Woloshin, 2007). As a result, patients can be misinformed and misled. An illustrative example is the Pill scare in the United Kingdom: when the U.K. Committee on Safety for Medicine stated that the risk of life-threatening blood clots in legs or lungs had increased twofold, that means by 100%, in the third generation of the oral contraceptive pill, many women stopped taking it. Results were unwanted pregnancies and abortions. In fact, the above-mentioned 100% can be translated as follows: 1 in 7000 women who took the second generation of the contraceptive pill suffered from blood clots; 2 in 7000 who took the third generation pill did (Gigerenzer & Gray, 2011). Communicating risk reductions and increases in relative rather than absolute formats leads to an overestimation of treatment effects (Covey, 2007).

Further evidence underlines the media's power to influence health attitudes, intentions, and behaviors. A survey by the National Health Council found that media reports led to health behavior change in 58% of the respondents (Roper Starch Worldwide, 1997). Even health professionals may be influenced by media reports when it comes to the prescription of treatments and drugs (Maclure et al., 1998). The Internet also affects the physician–patient relationship, as patients tend to form their beliefs and expectations before meeting the physician (Murray et al., 2003). In sum, information from the media becomes a double-edged sword: informed patients can increase the quality of the physician–patient interaction and improve health outcomes. However, misinformed patients might have increased or false expectations. They may also be influenced by a fundamental lack of the background

knowledge necessary for understanding health-relevant information (Galesic & Garcia-Retamero, 2010; Garcia-Retamero & Galesic, 2009; Schwartz, Woloshin, Black, & Welch, 1997). Interpreting health risks requires a minimum knowledge of statistical concepts and numeracy skills (Gigerenzer et al., 2007). The lack of background knowledge or an inability to deal with numbers can be addressed by framing the risks in the health environment in a certain way. For instance, the presentation of absolute risk reduction instead of relative risk reduction (Covey, 2007), the use of natural frequencies instead of conditional probabilities (Gigerenzer & Hoffrage, 1995), and the use of visual aids to communicate numerical information improve risk understanding substantially (Galesic, Garcia-Retamero, & Gigerenzer, 2009).

Intercultural comparison

Germany and Spain differ in their health systems, Internet usage, and HPV vaccine uptake. Spaniards are less proactive in seeking health information and have lower expectations of getting involved in health decisions (Delgado et al., 2010; Coulter & Jenkins, 2005). Computer and Internet use in Germany is more than twice the rate in Spain (WHO, 2011a), and a higher proportion of users search for health information in Germany (41%; Spain: 25%, see European Commission, 2010). The two countries differ on mortality rates of cervical cancer, with lower rates in Spain (WHO, 2011b). Finally, immunization rates differ: Whereas Germany reported a vaccination rate of 32% for girls aged 12–17 years in October 2009 (Fricke, 2010), Spain reported a rate of 77% for girls aged 11–14 years in the same period, achieved via school-based vaccination programs (Ministerio de Sanidad, 2011). In contrast, the German programs were opportunistic. Such cultural traditions can substantially affect the communication of information about HPV and the vaccine, as well as people's health-related behaviors.

In sum, the media have the power to change people's health attitudes and behaviors by influencing risk perception and the physician–patient relationship (Grilli, Ramsey, & Minozzi, 2009). It is therefore crucial to assess how the media communicate health (risk)

information to ensure transparent and balanced reporting. Biased reporting is evaluated on standards defining risk communication research, described below (Bodemer & Gaissmaier, 2012; Bunge, Mühlhauser, & Steckelberg, 2010; Gigerenzer et al., 2007, Steckelberg, Berger, Köpke, Heesen, & Mühlhauser, 2005).

Method

We conducted a two-step systematic literature search: the Internet search involved scanning websites from governmental institutions, health authorities, medical societies and associations, insurance providers, and pharmaceutical companies in Germany and Spain (see Appendix A). We used as search criteria "HPV" and "vaccination," or "human papillomavirus vaccination." We next performed a LexisNexis search identifying newspaper articles about HPV and cervical cancer in each country (see Appendix B). To document media reporting during the vaccine's implementation, search periods covered March 2007–June 2009 for newspaper reports and January 2009–May 2009 for websites. The Web search was restricted to the websites accessible during the search process.

Pre-defined inclusion criteria focused on reports that (a) intended to inform about the HPV vaccine, (b) had a minimum length of 200 words (to exclude brief notes or unspecific material about prevention programs), and (c) addressed primarily laypeople. To code the media reports, we reviewed medical literature on cervical cancer and the HPV vaccine (Neumeyer-Gromen, Bodemer, Müller, & Gigerenzer, 2011; see Appendix C). We focused on information important to decision makers when evaluating the HPV vaccine program for eventual participation. Our conceptual framework was based on the concept of "minimal statistical literacy" (Gigerenzer et al., 2007) and the criteria for evidence-based patient information (Bodemer & Gaissmaier, 2012; Bunge et al., 2010; Steckelberg et al., 2005). Gigerenzer et al. (2007) coined the term minimal statistical literacy to describe prerequisites for understanding (health) risks. The concept involves the acceptance of uncertainty in health, the adequate evaluation of risks (e.g., Does a risk refer to a lifetime risk? Does a risk refer to subpopulations? Does a risk refer to mortality or morbidity?), balanced information about

benefits and harms of a treatment, and the diagnostic value of a test (e.g., concepts like base rate, sensitivity, specificity, false alarm rates). Steckelberg et al. (2005) defined criteria for evidence-based patient information addressing transparency and comprehensibility. Criteria cover the content and the presentation format of information and suggest putting numerical information into context with other diseases or effects of alternative treatments. Bodemer and Gaissmaier (2012) discussed the advantages of numerical probabilities over verbal probability estimates, absolute over relative risk reductions, and natural frequencies over conditional probabilities, the benefits of visual aids, and transparent communication of uncertainties. Finally, Schwartz, Woloshin and Welch (2009) proposed fact boxes as a tool for summarizing medical evidence. Fact boxes include baseline risks, benefits, and harms of treatments by comparing a treatment group with a control group as well as information about etiology and the target population. Based on all above standards, the following dimensions for assessing reporting on websites and in newspapers were defined:

(1) Completeness (benefits and harms and side effects)

(2) Transparency (presentation of benefits and harms in absolute numbers instead of or at least in addition to relative numbers)

(3) Correctness (evidence-based information)

The evaluation criteria resulted in our coding scheme having three main content categories: (1) epidemiological, etiological, and pathological information about HPV, dysplasia, and cervical cancer, (2) information about the HPV vaccine and its benefits and harms, and (3) cost estimates of the vaccine. We applied the dimensions—completeness, transparency, and correctness—on each of the three categories. The fact box (Appendix C) displays the results of our literature search and provides background for the content categories. The scientific evidence is based on the findings for Gardasil in May 2009. Gardasil was the first vaccine approved and had the highest market share. Statistics about the vaccine's effectiveness were taken from the studies conducted to gain the vaccine's approval (see separate reference list for the fact box). To ensure comparability, the same coding

scheme was adopted for both media types. As LexisNexis does not map visualization, this category was limited to websites. Websites from health institutions and health authorities were only included if they were (1) characterized by high reliability and reputability and (2) easily accessible with common search engines such as Google, Yahoo, or Bing. We also included websites from the pharmaceutical companies producing the vaccine.

Two authors (ANG, NB) developed the coding scheme. It contained five sections: (1) identification of authors, communicators, and the target population; (2) epidemiological background with etiological and frequency information about cervical cancer; (3) evaluation of the balance of the report of effectiveness (i.e., discussion of pros and cons, concrete side effects and costs) and the information format (i.e., absolute or relative risk reduction measures); (4) the article's tone; and (5) visualization of the content. Additionally, general information about the media piece (e.g., information source, media type, date, and length) was coded. Three coders pretested the German coding scheme on a subsample of media reports. The revised coding scheme was then translated into Spanish by a bilingual speaker, pretested on a Spanish subsample and adapted to both countries. A second Spanish coder was trained by a bilingual German coder to assure equivalence of the coding processes. Three independent coders coded the German media reports, and two independent coders did the Spanish media pieces. In each group, 20% of the websites were rated to assure inter-rater reliability by Cohen's kappa, which revealed solid values between 0.6 and 0.7 (Grouven, Bender, Ziegler, & Lange, 2007).

Results

We summarize results for websites and newspapers separately and compare Germany and Spain within each media type. The three dimensions were translated into five subsections (see Table 1): communicators and targets, the epidemiological background, effectiveness of the intervention, the article's tone, and visualization (only for websites). The dimension completeness refers to balanced information about cervical cancer including pros and cons of getting vaccinated; the dimension transparency refers to the format used to explain

effectiveness by relative or absolute numbers, comparative figures and visualization; the dimension correctness refers to all the information evaluated by the first two dimensions. Data about effectiveness are of central interest. A total of 1586 and 2496 newspaper reports were identified in Germany and Spain, respectively. Of those, 141 and 293 articles met our inclusion criteria. For websites, 61 and 41 reports were identified.

Websites

Targets and communicators. In Germany, a majority of communicators to inform on the HPV vaccine was represented by scientists and doctors, whereas it was nearly exclusively communicated by "others" (i.e., governmental sources) in Spain. Differences between countries were also found for the target population, even though websites in both countries addressed mainly laypeople. German websites differentiated more between targets and showed high percentages targeted for parents (90%), relatives/peers (79%), girls or young women (84%), and doctors (66%). In contrast, only 39% of the Spanish websites directly addressed parents, and only 5% addressed girls or young women.

Epidemiological background. About two thirds of the websites provided numerical estimates of morbidity and mortality rates on cervical cancer (Germany: 66%; Spain: 68%); 57% of the German and 39% of the Spanish estimates were correct. Spanish websites reported causes of cancer (66%) and the possibility of spontaneous recovery (68%) more often than German websites (52% and 38%, respectively).

Effectiveness. Both countries' websites rarely mentioned dysplasia risk reduction. In Germany, 20% reported effectiveness in terms of relative risk reduction (10% gave correct estimates), and 5% as absolute risk reduction. In Spain, risk reduction was reported only in 5% of website reports (2% gave correct estimates). Reporting was similarly low for absolute risk reduction, prevention potential, and statements that the efficacy against cancer would be unclear. Cost estimates were rarely mentioned (Germany: 28%; Spain: 22%), only a few German websites compared the vaccination costs to cost estimates of other vaccines (8%). About two thirds of German websites explicitly recommended the vaccine (compared to 17%

of Spanish websites). In neither country did all the websites provide a reference to the Papanicolaou test for cancer screening (Germany: 61%; Spain: 73%). In general, German websites were more balanced (52% discussed both pros and cons; Spain: 37%), and half of them reported concrete side effects (30% as numerical estimates). One third of the Spanish websites included information about concrete side effects (all in numerical estimates), but predominantly as isolated positive proof of the vaccine's harmlessness as compared to other common vaccines (e.g., hepatitis).

Tone. We found two major differences between countries: Firstly, 41% of the Spanish websites compared with 64% of the German websites advertised the vaccine. Secondly, a majority of the Spanish websites (80%) primarily aimed at explaining the vaccine compared with only half of the German websites.

Visualization. Visualization was not very frequent on websites promoting the HPV vaccine. A few German websites (8%) provided visual material for the efficacy of the vaccine, but none of the Spanish websites did.

Newspapers

Targets and communicators. Looking at the newspapers, we found they were similar to the websites in terms of who was providing information. A majority of scientists and doctors represented communicators who informed about the vaccine in Germany (78% and 55%) in contrast with Spain (21% and 32%). In both countries, "others" also represented communicators in about half of all reports. Newspapers in both countries addressed mainly laypeople. German newspapers showed a greater variety in targets, resulting in high percentages for parents, relatives/peers, and girls or young women. Only 19% of Spanish newspapers addressed parents.

Epidemiological background. In both countries, about half of the newspaper reports (Germany: 54%; Spain: 45%) provided morbidity and mortality rates, and only half of these reports were correct in Spain (Germany: 43%; Spain: 20%). Both countries showed low percentages on the comparison to other diseases and spontaneous recovery (between 12% and

13%). Articles in Germany and Spain covered causes of cancer in 38% and 31% of the cases, respectively.

Effectiveness. Similar to websites, newspapers rarely communicated numbers for risk reduction in either country. Only German newspapers provided relative risk reduction (11%), with 6% providing correct numbers. Absolute risk reduction was only mentioned in 0.3% of the Spanish and 1% of the German articles; none of them were correct. Newspaper reports also lacked information about the prevention potential and statements that the effectiveness against cancer is unclear. Cost estimates were mentioned in 44% of the German and 66% of the Spanish newspapers with 6% of German and none of the Spanish newspapers comparing costs with cost estimates of other vaccines. One third of German newspapers but only 10% of Spanish newspapers explicitly recommended the vaccine. In both countries, about one third of the newspapers reported a reference to the Papanicolaou test for cancer screening. Reporting was more balanced in Germany than in Spain (i.e., in Germany, half of the newspapers reported pros and cons, in Spain only 17% did so). No cross-cultural differences emerged on the communication of concrete side effects (Germany: 14%; Spain: 11%). Still, newspapers in both countries provided less information than websites on most key aspects, such as baseline risk, cancer causes, spontaneous recovery, effectiveness, and side effects.

Tone. Of the German newspaper reports, 28% advertised the vaccination, whereas 17% did so in Spain. In general, reports in both countries aimed at explaining the vaccine or vaccination programs (72% for both countries).

Table 1

Comparison of Internet and Newspaper Reporting on the HPV Vaccine in Germany and Spain.

	Website reports		Newspaper reports	
	Germany (*N*=61) % (*n*)	Spain (*N*=41) % (*n*)	Germany (*N*=141) % (*n*)	Spain (*N*=293) % (*n*)
Communicators[a]				
Scientists	84% (51)	10% (4)	78% (110)	21% (62)
Doctors	70% (43)	39% (16)	55% (77)	32% (95)
Patient representatives	15% (9)	2% (1)	8% (11)	0%
Girls or young women[b]	46% (28)	5% (2)	6% (9)	0%
Other[c]	51% (31)	80% (33)[c]	51% (72)	44% (130)
Target group[a]				
Laypeople in general	90% (55)	88% (36)	99% (139)	97% (284)
Doctors	66% (40)	41% (17)	12% (17)	1% (4)
Scientists	54% (33)	15% (6)	12% (17)	1% (4)
Parents	90% (55)	39% (16)	71% (100)	19% (56)
Relatives/Peers	79% (48)	22% (9)	56% (79)	1% (3)
Target population/girls or young women	84% (51)	5% (2)	69% (97)	0.3% (1)
Information centre	54% (33)	0%	12% (17)	0%
Other	56% (34)	0%	46% (65)	0%
Epidemiology of Cervical Cancer				
Frequencies	66% (40)	68% (28)	54% (76)	45% (133)
Correct estimates	57% (35)	39% (16)	43% (61)	20% (60)[d]
Comparison to other diseases	30% (18)	0%	12% (17)	13% (39)
Causes of cancer	52% (32)	66% (27)	38% (54)	31% (90)
Spontaneous recovery	38% (23)	68% (28)	13% (18)	12% (36)
Effectiveness				
RRR dysplasia	20% (12)	5% (2)	11% (15)	0%
Correct estimates	10% (6)	2% (1)	6% (8)	0%
ARR dysplasia	5% (3)	0%	1% (2)	0.3% (1)
Correct estimates	0%	0%	0%	0%
Prevention potential	2% (1)	2% (1)	0%	1% (4)
Correct estimates	2% (1)	2% (1)	0%	0%
Efficacy against cancer unclear	13% (8)	0%	0%	0%
Cost estimates	28% (17)	22% (9)	44% (62)	66% (193)
Comparison to other cost estimates	8% (5)	0%	6% (8)	0%
Explicit recommendation to get vaccinated	66% (40)	17% (7)	29% (41)	10% (30)

(continued)

Table 1.

Comparison of Internet and Newspaper Reporting on the HPV Vaccine in Germany and Spain (continued).

	Website reports		Newspaper reports	
	Germany (*N*=61) % (*n*)	Spain (*N*=41) % (*n*)	Germany (*N*=141) % (*n*)	Spain (*N*=293) % (*n*)
Reference to Pap screening	61% (37)	73% (30)	36% (51)	33% (97)
Pros and cons	52% (32)	37% (15)	50% (70)	17% (49)
Concrete side effects	48% (29)	32% (13)	14% (20)	11% (33)
Numerical estimates	30% (18)	32% (13)	7% (10)	5% (16)
Article's tone[a]				
Advertising	64% (39)	41% (17)	28% (39)	17% (50)
Explaining	51% (31)	80% (33)	72% (101)	72% (210)
Warning	3% (2)	0%	5% (7)	14% (21)
Other	3% (2)	5% (2)	4% (5)	39% (113)
Visualization[e]				
Epidemiology/Frequency	10% (6)	5% (2)		
Effectiveness	8% (5)	0%		
Biology/Virus/Anatomy	11% (7)	17% (7)		
Pictures (i.e. photos)	21% (13)	29% (12)		

Abbreviations: RRR, relative risk reduction; ARR, absolute risk reduction; Pap, Papanicolaou test.
[a] Multiple rankings for these categories were possible.
[b] The category "girls and or young women" refers to testimonials of this group only and not to self-reports or blog contents.
[c] The high percentage in the category "others" reflects the amount of governmental communicators in websites that provided information about the vaccine.
[d] This difference is in part due to the fact that Spanish newspapers often reported numbers without reference to the base rate and age standardization of the specific Spanish communities or the country of Spain, respectively.
[e] As LexisNexis does not allow to map visualization only websites were evaluated in this category.

Discussion

We sought to evaluate media coverage of the HPV vaccine. We compared two media sources—Internet websites and newspapers—in two countries with different health systems—Germany and Spain. The three evaluation dimensions—completeness, transparency, and correctness—revealed shortcomings in both countries and both media types. In Germany, about half of the websites communicated correct epidemiological information and causes of cancer, but only one third mentioned spontaneous recovery. In Spain, websites mentioned

spontaneous recovery more often, but epidemiological information was less frequently correct. In general, newspapers documented epidemiological, etiological, and pathological information less often than websites. According to our evaluation criteria, both media types in both countries failed to provide correct and transparent information about effectiveness—with Germany showing slightly higher proportions of correctness and transparency.

Despite the limited evidence regarding the vaccine's evaluation (e.g., duration of protection1), only German websites (13%) explicitly referred to the scientific uncertainty of the vaccine, stating that efficacy against cancer is still unclear. Half of the websites and newspaper reports discussed pros and cons in Germany; this was less common in Spanish reports (37% and 17%, respectively). Websites reported side effects more often than newspapers. In both countries, only one third of newspaper reports referred to the Papanicolaou test, but the majority of websites did so. On the other hand, newspaper reports in both countries provided cost estimates more frequently than websites. Target groups of most reports were laypeople. While the majority of reports had an expository tone, more than half of the German websites had an advertising tone and two thirds explicitly recommended the vaccine. Websites in both countries used visualization rarely, and if so, illustrations seldom seem intended to facilitate understanding of risk information (e.g., icon arrays).

Although differences between media types were not analyzed, potential disparities may be due to the target population: Websites may inform girls and parents and therefore include epidemiological and etiological information, side effects, and references to Papanicolaou screening. Newspapers may target the broader public, providing less background information but discussing cost estimates more frequently. The greater number of governmental and ministerial communicators on Spanish websites may be due to differences between the health care systems: the national health care system in Spain is centrally organized, with systematic school-based vaccination programs. In Germany, the system is organized in a more decentralized, self-administered way and vaccination is offered opportunistically. This communication style may lead to a more directive, less informative,

and, finally, less participative health care system. Fewer demands for and less active interest in transparent, balanced media reporting by Spanish citizens could be a result. Future research should address these hypotheses in more detail and further investigate the effects of different health systems on public health decisions.

These descriptive results may be limited by a rather hypothesis-generating than hypothesis-testing approach for identifying differences in media coverage between media sources and countries. The LexisNexis search and the identification criteria for websites, however, reflect a representative sample of current media coverage. Differences between media types may have occurred because of the different timeframes of our literature search (newspapers: March 2007–June 2009; websites: January 2009–May 2009), but no differences were found in newspaper coverage before and after general vaccine criticism arose in 2008 (Dören et al., 2008), 2009 (Gerhardus, Dören, & Gerlach, 2009), or 2010 (Gerhardus & Razum, 2010). We intentionally restricted our Internet search to health authorities, since public health websites and university websites scored highest on credibility, accessibility, content, and design (Tozzi et al., 2010).

As noted earlier, the media can influence health attitudes and behavior. While health communications often are intended to persuade or nudge people (Marteau, Ogilvie, Roland, Suhrcke, & Kelly, 2011), an alternative communication strategy takes an informative, nondirective approach that centers on transparency and (gained) knowledge (Feufel & Bodemer, 2012). The goal is to equip patients with sufficient knowledge to make individual, informed decisions—the basis for informed consent and shared decision making. The media can be one source for independent and transparent health information. In line with other researchers, we suggest that fact boxes can provide all the relevant information (Schwartz et al., 2009). One might argue that the fact box presented here is too complex and detailed—but it allows each individual to select the information needed to make a personal decision. Furthermore, health professionals and journalists can benefit from facts boxes to communicate key aspects of treatment effectiveness and shortcomings.

To improve future media coverage, reporting standards—such as CONSORT (2011) or STROBE (2011)—should be developed, with equal access for journalists, public health policy makers, health care professionals, and citizens. Standards will help consumers identify reliable and balanced information sources and will support the use of transparent formats to translate scientific knowledge. Based on current research about risk communication and concepts of minimal statistical literacy and criteria for evidence-based patient information, we propose standards for media coverage (Figure 1). Similar to the International Patient Decision Aid Standards (Elwyn et al., 2006; Holmes-Rovner, 2007), our standards may help further people's involvement in and knowledge about health decisions (O'Connor et al., 2001).

Standards for media coverage of health issues

- Explain goal of medical treatment or prevention (i.e., primary vs. secondary prevention programs)
- Define target population
- Explain etiology (i.e., causes of cancer, possibility of spontaneous recovery) so that underlying hypotheses for the development of new treatment or prevention options can be understood
- Provide epidemiological data (i.e., number of incidences and number of deaths) to convey an idea of baseline risk
- Communicate treatment effects in absolute numbers (i.e., absolute risk reduction). Additionally, communicate side effects in absolute numbers (i.e., risk increase, false positives) to convey information for shared decision making
- Include visualizations to illustrate treatment effects to address a wide range of people (also those with limited numeracy skills)
- Communicate cost estimates to convey an idea of individual and public health investments in view of limited individual public resources
- Mention alternatives to treatment
- Use comparative figures (i.e., effects of other well-established prevention programs, overall cancer mortality, costs of other well-known interventions, etc.)
- Disclose uncertainties and what is not (yet) known (i.e., duration of immunization, effects on cancer since only surrogate measures are available, etc.)
- Disclose conflicts of interests

Figure 1. Proposed standards for transparent online and offline information about health interventions.

Author Note

Conflict of interest statement: None of the authors had any financial interests or conflicts of interests. Funding: This research was supported by the joint program "Acciones Integradas HispanoAlemanas" from the DAAD and the Ministerio de Ciencia y Tecnología and by the projects "How to Improve Understanding of Risks about Health (PSI2008-02019)" funded by the Ministerio de Ciencia e Innovación (Spain), "Helping Doctors and Their Patients Make Decisions about Health (PSI2011-22954)" funded by the Ministerio de Economía y Competitividad (Spain), and the joint program "Strategies to Improve Comprehensions of Risks about Health" (Proyecto de Movilidad; Acciones Integradas), funded by the DAAD (Germany) and the Ministerio de Ciencia y Tecnología (Spain). The ethics committee of the Max Planck Institute for Human Development approved the research. We thank Anita Todd for editing the paper.

References

Abdelmutti, N., & Hoffman-Goetz, L. (2009). Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: A content analysis of Canadian and U.S. national newspaper articles. *Women & Health*, *49*(5), 422–440. doi:10.1080/03630240903238776

Bodemer, N., & Gaissmaier, W. (2012). Risk communication in Health. In R. Hillerbrand, P. Sandin, S. Roeser & M. Peterson (Eds.). *Handbook of Risk Theory* (pp. 621–660). Springer Netherlands.

Chapman, S., & Lupton, D. (1994). *The fight for public health: Principles and practice of media accuracy*. London: British Medical Journal.

Bunge, M., Mühlhauser, I., & Steckelberg, A. (2010). What constitutes evidence-based patient information? Overview of discussed criteria. *Patient Education and Counseling*, *78*(3), 316–328. doi:10.1016/j.pec.2009.10.029

CONSORT (2011). *Consolidated standards of reporting trials* [cited 2011 Feb 1]. Retrieved from: http://www.consort-statement.org/consort-statement/

Cotten, S. R., & Gupta, S. S. (2004). Characteristics of online and offline health information seekers and factors that discriminate between them. *Social Science & Medicine*, *59*(9), 1795–1806. doi:10.1016/j.socscimed.2004.02.020

Coulter, A., Jenkins, C. (2005) European patients' views on the responsiveness of health systems and healthcare providers. *European Journal of Public Health*, 15, 355–60.

Covey, J. (2007). A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making*, *27*(5), 638–654. doi:10.1177/0272989X07306783

D'Alessandro, D. M., Kreiter, C. D., Kinzer, S. L., & Peterson, M. W. (2004). A randomized controlled trial of an information prescription for pediatric patient education on the Internet. *Archives of Pediatrics & Adolescent Medicine*, *158*(9), 857–862. doi:10.1001/archpedi.158.9.857

Delgado, A., López-Fernández, L. A., Juan de Dios, L., Saletti Cuesta, L., Gil Garrido, N., Puga González, A. (2010). *Patient expectation of decision making for distinct health problems*.

[Internet]. 2010 Jan [cited 2011 Feb 1] Retrieved from: http://www.scielosp.org/scielo.php?pid=S0213-91112010000100011&script=sci_arttext

Dören, M., Gerhardus, A., Gerlach, F.M., Hornberg, C., Kochen, N.M., Kolip, P., et al. (2008). *Wissenschaftlier/innen fordern Neubewertung der HPV-Impfung und ein Ende der irreführenden Informationen* [Internet]. 2008 Nov [cited 2011 Feb 1]. Retrieved from: http://www.uni-bielefeld.de/gesundhw/ag3/downloads/Stellungnahme_Wirksamkeit_HPV-Impfung.pdf.

Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., Thomson, R., et al. (2006). Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *British Medical Journal*, *333*(7565), 417. doi:10.1136/bmj.38926.629329.AE

European Commission. (2011). *Europe's Digital Competitiveness Report: Main achievements of the i2010 strategy 2005-2009*. [Internet]. 2011 [cited 2011 Oct 1]. Retrieved from: http://ec.europa.eu/information_society/eeurope/i2010/docs/annual_report/2009/digital_competitiveness.pdf

Feufel, M.A. & Bodemer, N. (2012). *Finding the right tool to improve health decisions: Nudging, social marketing, empowerment.* Manuscript in preparation.

Fox, S. (2011). Health Topics: 80% of internet users look for health information online. [Internet]. 2011 [cited 2011 Oct 1]. Retrieved from: http://pewinternet.org/Reports/2011/HealthTopics.aspx

Fricke A. (2010). *Langfristiger Erfolg der HPV Impfung erwartet.* Ärztezeitung [Internet]. [cited 2011 Feb 1]. Retrieved from: http://www.aerztezeitung.de/politik_gesellschaft/article/592205/langfristiger-erfolg-hpv-impfung-erwartet.html

Frost, K., Frank, E., Maibach, E. (1997). Relative risk in the news media: A quantification of misrepresentation. *American Journal of Public Health*, 87,842–845.

Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, *170*(5), 462–468. doi:10.1001/archinternmed.2009.481

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American Journal of Public Health*, *99*(12), 2196–2202. doi:10.2105/AJPH.2009.160234

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210–216. doi:10.1037/a0014474

Gerhardus, A., Dören, M., Gerlach, F.M. (2009) Gebährmutterhalskrebs: Wie wirksam ist die HPV-Impfung? *Deutsches Ärzteblatt*, 106, 330–334.

Gerhardus, A., & Razum, O.(2010). A long story made too short: Surrogate variables and the communication of HPV vaccine trial results. *Journal Epidemiological Community Health, 64*, 377–378.

Gigerenzer, G., & Gray, J.A.M. (2011). Launching the century of the patient. In: G. Gigerenzer & J.A.M. Gray (Eds.), *Better doctors, better patients, better decisions: Envisioning health care 2020* (pp. 3–28). Cambridge, MA: MIT Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Gigerenzer, G., Gaissmaier, W.**,** Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8,* 53–96.

Gigerenzer, G., Wegwarth, O., Feufel, M.A. (2011). Misleading communication of risk. *British Medical Journal*, c4830.

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health information-seeking behaviour in adolescence: The place of the internet. *Social Science & Medicine (1982)*, *60*(7), 1467–1478. doi:10.1016/j.socscimed.2004.08.010

Grilli, R., Ramsey, C., Minozzi, S. (2009). Mass media interventions: Effects on health care utilization (Review). *Cochrane Database Systematic Review*, 1:CD000389.

Habel, M. A., Liddon, N., & Stryker, J. E. (2009). The HPV vaccine: A content analysis of online news stories. *Journal of Women's Health (2002)*, *18*(3), 401–407. doi:10.1089/jwh.2008.0920

Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., & Viswanath, K. (2005). Trust and sources of health information: the impact of the Internet and its implications for health care providers: Findings from the first Health Information National Trends Survey. *Archives of Internal Medicine*, *165*(22), 2618–2624. doi:10.1001/archinte.165.22.2618

Holmes-Rovner, M. (2007). International Patient Decision Aid Standards (IPDAS): Beyond decision aids to usual design of patient education materials. *Health Expectations*, *10*(2), 103–107. doi:10.1111/j.1369-7625.2007.00445.x

James, C., James, N., Davies, D., Harvey, P., & Tweddle, S. (1999). Preferences for different sources of information about cancer. *Patient Education and Counseling*, *37*(3), 273–282.

Kelly, B. J., Leader, A. E., Mittermaier, D. J., Hornik, R. C., & Cappella, J. N. (2009). The HPV vaccine and the media: how has the topic been covered and what are the effects on knowledge about the virus and cervical cancer? *Patient Education and Counseling*, *77*(2), 308–313. doi:10.1016/j.pec.2009.03.018

Khoo, K., Bolt, P., Babl, F. E., Jury, S., & Goldman, R. D. (2008). Health information seeking by parents in the Internet age. *Journal of Pediatrics and Child Health*, *44*, 419–423. doi:10.1111/j.1440-1754.2008.01322.x

Kurzenhäuser, S. (2003) Welche Informationen vermitteln deutsche Gesundheitsbroschüren über die Screening-Mammographie? *Zeitschrift für Ärtzliche Fortbildung und Qualitätssicherung*, 97, 53–57.

Maclure, M., Dormuth, C., Naumann, T., McCormack, J., Rangno, R., Whiteside, C., & Wright, J. M. (1998). Influences of educational interventions and adverse news about calcium-

channel blockers on first-line prescribing of antihypertensive drugs to elderly people in British Columbia. *Lancet*, *352*(9132), 943–948. doi:10.1016/S0140-6736(97)11390-3

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, *8*(10), 543–548.

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: Can nudging improve population health? *British Medical Journal*, *342*, d228.

Martin-Llaguno, M., & Alvarez-Dardet, C. (2010). *Withdrawal of an advertising campaign to promote the quadrivalent human papilloma virus vaccine in Spain. Gac Sanit* [Internet]. 2010 Jan [cited 2011 Feb 1]; 24(1):75–7. Retrieved from: http://www.scielosp.org/scielo.php?pid=S0213-91112010000100013&script=sci_arttext

Morahan-Martin, J. M. (2004). How internet users find, evaluate, and use online health information: a cross-cultural review. *Cyberpsychology & Behavior*, *7*(5), 497–510.

Moynihan, R., Bero, L., Ross-Degnan, D., Henry, D., Lee, K., Watkins, J., Mah, C., et al. (2000). Coverage by the news media of the benefits and risks of medications. *The New England Journal of Medicine*, *342*(22), 1645–1650. doi:10.1056/NEJM200006013422206

Ministerio de Sanidad, Politica Social e Igualidad (2011). *Coberturas de Vacunación* [cited 2011 Feb 1]. Retrieved from: http://www.msps.es/profesionales/saludPublic/prevPromocion/vacunacion/coberturas.htm

Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., White, M., Zapert, K., et al. (2003). The impact of health information on the internet on the physician-patient relationship: patient perceptions. *Archives of Internal Medicine*, *163*(14), 1727–1734. doi:10.1001/archinte.163.14.1727

Neumeyer-Gromen, A., Bodemer, N., Müller, S. M., & Gigerenzer, G. (2011). Ermöglichen Medienberichte und Broschüren informierte Entscheidungen zur Gebärmutterhalskrebsprävention? *Bundesgesundheitsblatt*, *54*, 1197–1210. doi:10.1007/s00103-011-1347-5

O'Connor, A. M., Stacey, D., Rovner, D., Holmes-Rovner, M., Tetroe, J., Llewellyn-Thomas, H., Entwistle, V., et al. (2001). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews (Online)*, (3), CD001431. doi:10.1002/14651858.CD001431

Roper Starch Worldwide (1997). *Americans Talk about Science and Medical News: The National Health Council Report*. Prepared for the National Health Council by Roper Starch Worldwide New York, NY.

Sarfati, D., Howden-Chapman, P., Woodward, A., & Salmond, C. (1998). Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. *Journal of Medical Screening*, *5*(3), 137–140.

Schönbach K., de Waal, E., Lauf, E. (2005) Online and print newspapers: Their impact on the extent of the perceived public agenda. European Journal of Communication;20:245–258.

Schwartz, L.M., Woloshin, S., Black, W.C., Welch, H.G. (1997) The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127,966–72.

Schwartz, L. M., Woloshin, S., & Welch, H. G. (2009). Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Annals of Internal Medicine*, *150*(8), 516–527.

Sedrakyan, A., & Shih, C. (2007). Improving depiction of benefits and harms: Analyses of studies of well-known therapeutics and review of high-impact medical journals. *Medical Care*, *45*(10 Supl 2), S23–28. doi:10.1097/MLR.0b013e3180642f69

Slater, M. D., Long, M., Bettinghaus, E. P., & Reineke, J. B. (2008). News coverage of cancer in the United States: a national sample of newspapers, television, and magazines. *Journal of Health Communication*, *13*(6), 523–537. doi:10.1080/10810730802279571

Steckelberg, A., Balgenorth, A., & Mühlhauser, I. (2001). Analyse von deutschsprachigen Verbraucher-Informationsbroschüren zum Screening auf kolorektales Karzinom. *Zeitschrift für Ärtzliche Fortbildung und Qualitätssicherung,* 95,535–538.

Steckelberg, A., Berger, B., Köpke, S., Heesen, C., Mühlhauser, I. (2005). Criteria for evidence-based patient information. *Zeitschrift für Ärtzliche Fortbildung und Qualitätssicherung*, 99: 343–351.

Strobe Statement (2011). *Strengthening the reporting of observational studies in epidemiology* [cited 2011 Feb 1]. Retrieved from: http://www.strobe-statement.org

Stryker, J. E., Moriarty, C. M., & Jensen, J. D. (2008). Effects of newspaper coverage on public knowledge about modifiable cancer risks. *Health Communication*, *23*(4), 380–390. doi:10.1080/10410230802229894

Tozzi, A. E., Buonuomo, P. S., Ciofi degli Atti, M. L., Carloni, E., Meloni, M., & Gamba, F. (2010). Comparison of quality of internet pages on human papillomavirus immunization in Italian and in English. *The Journal of Adolescent Health*, *46*(1), 83–89. doi:10.1016/j.jadohealth.2009.05.006

Tuffrey, C., & Finlay, F. (2002). Use of the internet by parents of paediatric outpatients. *Archives of Disease in Childhood*, *87*(6), 534–536.

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Education & Behavior*, *27*(5), 591–615.

World Health Organization (2011a). *Core health indicators* [cited 2011 Feb 1]. Available from: http://apps.who.int/whosis/database/core/core_select_process.cfm?countries=all&indicators=nha

World Health Organization (2011b). *Preparing for the introduction of HPV vaccine in the WHO European region* [cited 2011 Feb 1]. Retrieved from: http://www.euro.who.int/__data/assets/pdf_file/0007/98746/E91432.pdf

Zimmerman, R. K., Wolfe, R. M., Fox, D. E., Fox, J. R., Nowalk, M. P., Troy, J. A., & Sharp, L. K. (2005). Vaccine criticism on the World Wide Web. *Journal of Medical Internet Research*, *7*(2), e17. doi:10.2196/jmir.7.2.e17

Appendix A

German institutions include the following: *Bundesministeriums für Gesundheit (BMG); Bundeszentrale für gesundheitliche Aufklärung; Robert-Koch-Institut; Paul-Ehrlich-Institut; Gemeinsamer Bundesausschuss; Krankenkassen und deren Bundesverbände; Bundesärztekammer und alle Landesärztekammern; Landesgesundheitsministerien und Landesgesundheitsämter; Bundesbeauftragte der Bundesregierung für die Belange der Patientinnen und Patienten; Deutsches Krebsforschungsinstitut; Gynäkologische und Pädiatrische Fachgesellschafte; Deutsche Gesellschaft für Epidemiologie; Deutsches Netzwerk Evidenzbasierte Medizin; Deutsche Gesellschaft für Sozialmedizin und Prävention; Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie; Krebsfachgesellschaften, Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften; Pro Familia; Selbsthilfe Kontakt und Information Stelle (SEKIS); Nationale Kontakt- und Informationsstelle zur Anregung und Unterstützung von Selbsthilfegruppen (NAKOS); "www.patienteninformation.de" von Bundesärztekammer und Kassenärztlicher Bundesvereinigung; "www.akdae.de" der Arzneimittelkommission der deutschen Ärzteschaft, Techniker Krankenkasse und BMG; "*www.gesundheitsinformation.de*" Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; Sanofi Pasteur; Merck & Co.; "*www.zervita.de*" (Zervita)*

Spanish institutions include the following: *Accudes; Adeslas; Ampligen; Asisa; Asociación Española Contra el Cáncer; Asociación Española de Patología Cervical y Colposcopia; Comunidad Castilla de la Mancha; Comunidad Foral de Navarra; Comunidad de Madrid Consejería de Sanidad; Comunidad de Madrid; Consellería de Sanitat Valencia; DVK Seguros; Fisterra- Lorena Anido Redondo (GP); Generalitat de Catalunia; GlaxoSmithKline; Gobierno de Aragón; Gobierno de Canarias; Gobierno de Asturias; Gobierno de Murcia; Govern de Illes Balears; Gobierno de la Rioja; Gobierno Salud de la Rioja; Gobierno Vasco;*

Junta Castilla y León; Junta de Andalucía; Junta de Galicia; Labec Pharma; Mapfre Seguros; Ministerio de Sanidad y Política Social; Sanidad Asturiana; Sanidad Murciana; Sanidad Vasca; Sanitas; Sanofi Pasteur MSD; Sociedad Aragonesa de Medicina Familiar y Comunitaria; Sociedad Española de Contracepción; Sociedad Española de Ginecología y Obstetricia; Sociedad Española de Medicina General; Sociedad Española de Medicina Familiar y Comunitaria; Sociedad Española de Medicina Rural y Generalista; Asociación Española de Pediatría de Atención Primaria; Sociedad Española de Salud Pública y Administración Sanitaria; Sociedad Navarra de Medicina de Familia y Atención Primaria; Sociedad de Pediatría de Atención Primaria de Extremadura; Societat Valenciana de Medicina Familiar y Comunitaria

Appendix B

	Search string for LexisNexis search on HPV vaccination
German	(hpv! OR papillom! OR kondylom! OR condylom! OR dna-vir! OR hp-vir! OR genitalwarz! OR warzenvir! OR gebärmutterhalsk! OR gebärmutterhalst!) AND (impf! OR vakzin! OR immunisierung OR protektiv OR protektion OR prävention OR vorbeugung OR gardasil OR sanofi pasteur OR cervarix OR zervarix OR glaxo smithkline OR silgard OR merck)
Spanish	(hpv! OR papiloma humano! OR condiloma! OR ph-vir! OR verrugas en el pene! OR condiloma acuminata! OR verrugas venéreas! OR cáncer cervical! OR cuello de útero! OR cáncer de cuello uterino! OR neoplasia cervical intraepitelial! OR) AND (vacuna! OR papanicolaou! OR frotis de pap! OR inmunización OR proteción OR prevención OR gardasil OR sanofi pasteur OR cervarix OR glaxo smithkline OR silgard OR merck)

Appendix C

Facts box summarizing evidence for information about cervical cancer and the HPV vaccine (Gardasil). The facts box is based on scientific evidence (May 2009) and served as baseline comparison for the evaluation of media reports of German and Spanish websites and newspapers. A coding scheme was developed including all information displayed in the facts box.

Human papillomavirus (HPV) vaccination with Gardasil

What is the aim of the vaccine?[1,2]	Preventing the infection with HPV type 16 and 18 to decrease the risk of cervical cancer; additional protection against genital warts
How is HPV transmitted?[1,2]	By sexual contact
What are the consequences of an infection?[3-6]	Infections with potentially 18 different types of HPV over decades can lead to changes in the tissue, which may (1) cause preliminary stages of cancer in the cervix, which may develop into (2) cervical cancer. Seventy of 100 cases of cervical cancer are due to HPV 16/18.

How prevalent is cervical cancer?

		In 100,000 women per year		*In all women per year*	
		Germany[7]	Spain[8]	Germany[7]	Spain[8]
Deaths	Cervical cancer	*3*	*2.5*	1,500	718
	All types of cancer	*230*	*236*	*101,000*	*6,565*
Incidence	Cervical cancer	*15*	*7.6*	6,200	1,965
	All types of cancer	*500*	*450*	*200,000*	*121,176*

Is there a chance that the infection will disappear without treatment?[3,5,6]	Yes. For orientation, there is spontaneous recovery in over 90 of 100 cases for infections and in 50 of 100 cases for preliminary stages of cancer.
For whom is it recommended *and* covered by public insurance?[1,2]	Girls 12–17 years (Germany) and 11–14 years (Spain), preferably before any first sexual contact. Some Spanish communities also recommend vaccination for girls 9–15 years.
How long does the vaccine last?[1,2]	Minimum 5 years.
Are there other types of HPV that could increase and/or decrease after the vaccination?[1]	Due to theoretical assumptions potentially yes. This is called replacement and cross protection.
Are there other preventive methods?[1,9]	Yes. Early detection with Papanicolaou/Pap test "for woman aged 20 years and older" (Germany) or "between 15 and 25 until 49–65 years" (Spain, differing by community), which should also be

applied for vaccinated women. Use of condoms.

How effective was Gardasil in scientific studies?[10-12]*

	Of 1,000 women	
Incidence of risky, preliminary stages of cervical cancer (due to all HPV viruses)	*Vaccinated*	*Not vaccinated*
At the beginning of the study, all participants had not been infected with the types of HPV that are covered by the vaccine/all were virgins[10]	20	28
All participants (at the beginning of the study infection with HPV possible)[10-12]	42	49
Incidence of cervical cancer[12]	Not clear, no scientific evidence	

Are there side effects of Gardasil? (based on the European release)[12-14]

Very frequent – frequent	*Occasional – rare*	*Very rare*
>1,000 – ≥10,000 of 100,000	*10 – 1,000 of 100,000*	*< 10 of 100,000*
- fever; injection site: redness, pain, swelling, effusion, itching	- unspecific arthritis, joint trouble - severe allergic reaction, urticaria	- bronchoconstriction with severe shortness of breath

On the basis of spontaneous reports after the release of the vaccine (size and estimated number of unreported cases is unclear). These reports are in temporal relation to the vaccine; it is unclear whether the vaccine caused these incidences.

- serious neurological illness (Guillain-Barré-Syndrom), signs of paralysis, paralysis of the face, seizure
- vomiting, muscle pain, lymphadenopathy, allergic reaction
- sporadic cases of death

What are the vaccination costs?

	Germany[15-17]	*Spain*[18,19]
Costs for one complete vaccination	*465 euros*	*465 euros*
Total cost for one cohort of girls	*about 200 million euros*	*about 63 million euros*
Total cost of all annual public health programs (for all diseases)	*about 1883 million euros*	*About 946 million euros*

1. Robert Koch Institut/ RKI (2007) Mitteilung der Ständigen Impfkommission/ STIKO am RKI (Stand: März 2007). Epid Bull 12: 97-103
2. Robert Koch Institut/ RKI (2010): Empfehlungen der Ständigen Impfkommission/ STIKO am RKI/Stand Juli 2010. Epid Bull 30: 235–250
3. Wentzensen N, Klug S J (2009) Früherkennung des Zervixkarzinoms: Suche nach einem Gesamtkonzept. Dtsch Arztebl 105: 617-622

4. Munoz N, Bosch X, de Sanjose S et al. (2003) Epidemiologic classification of Human Papillomavirus Types associated with cervical cancer. N Engl J Med 348(6): 518-527
5. Schiffman M (2007) Integration of human papillomavirus vaccination, cytology, and human papillomavirus testing. Cancer 111:145–53
6. Ostor A G (1993) Natural history of cervical intraepithelial neoplasia: a critical review. Int J Gynecol Pathol 12:186–192
7. Statistisches Bundesamt (2009) Gesundheitsberichterstattung des Bundes (2009). http://www.destatis.de
8. Ferlay J, Parkin D M, Steliarova-Foucher E.. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* 2010;46:765–81.
9. Winer R L, Hughes J P, Feng Q, O'Reilly S, Kiviat M B, Holmes K K, Koutsky L A (2006) Condom use and the risk of genital human Papillomavirus infection in young women. N Engl J Med, 354(25): 2645-2654
10. FUTURE II Study Grp (2007) Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N. Engl. J. Med. 356: 1915-1927
11. FUTURE I investigators (2007) Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N. Engl. J. Med. 356:1928-1943
12. European Medicine Agency/ EMA (2008) Europ. Beurteilungsbericht (EPAR) GARDASIL (31/10/2008 Gardasil-H-C-703- II-13). http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000703/human_med_000805.jsp&mid=WC0b01ac058001d125&murl=menus/medicines/medicines.jsp und http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Procedural_steps_taken_and_scientific_information_after_authorisation/human/000703/WC500021147.pdf
13. Centers for Disease Control and Prevention (CDC) (2008) Vaccine Adverse Event Reporting System (VAERS) 2008. http://www.vaers.hhs.gov/scripts/data.cfm
14. European Medicine Agency/ EMA (2010): Gardasil: EPAR – Summary for the public (14/09/2010) http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Summary_for_the_public/human/000703/WC500021146.pdf und http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Procedural_steps_taken_and_scientific_information_after_authorisation/human/000703/WC500021147.pdf
15. a-t (Arznei-Telegramm) (2008) 1997 und 2007 im Vergleich - die Umsatzstärksten Arzneimittel. a-t 39:65-66
16. Rosenbrock R (2007) HPV-Impfung - Durchbruch in der Krebsprävention? http://www.forum-gesundheitspolitik.de/artikel/artikel.pl?artikel=0644 oder http://www.forum-gesundheitspolitik.de/dossier/PDF/Rosenbrock-HPV-Impfung.pdf
17. OECD (2011) Dataset System of Health Accounts, Country Germany http://www.oecd.org/document/23/0,3746,en_33873108_33873402_33884311_1_1_1_1,00.html
18. OECD (2011) Dataset Dataset System of health Accounts, Country Spain http://www.oecd.org/document/17/0,3343,en_2649_33929_33884753_1_1_1_1,00.html
19. Lopez Alemany J M, Cortes Bordoy J, Gil de Miguel A. Human papilomavirus tetravalent vaccine cost-effectiveness study. *Rev Esp Econ Salud* 2007;6: 400-8.

*The numbers of the first effectiveness data can be found on page 1922, 2nd paragraph, in the Future II-study; this result corresponds to the a-priori defined „unrestricted susceptible population" which is similar to the combined analyses of the Future I- and Future II-studies of the "modified intention-to-treat-analyses" (MITT-2) in the EMA report 2008 on page 22 in table 15.[12] The numbers of the 2nd effectiveness data can be found on page 22, last paragraph, and in table 15 as the a-priori defined "MITT-3"-analysis; it is also a combined analysis of the Future I- and Future II-studies.[12]

Chapter 4

Improving the understanding of relative risk information with baseline risk: The role of presentation format and numeracy

Nicolai Bodemer[1, 2], Björn Meder[2], & Gerd Gigerenzer[1, 2]

[1]Harding Center for Risk Literacy, Berlin, Germany
[2]Max Planck Institute for Human Development, Berlin, Germany

[Word count: 7525]

Bodemer, N., Meder, B., & Gigerenzer, G. (2012). *Improving the understanding of relative risk information with baseline risk: The role of presentation format and numeracy*. Manuscript submitted for publication.

Abstract

Relative risk reductions, a standard format for communicating treatment benefits in medical journals, decision aids, and patient brochures, are often misunderstood by both doctors and patients. To resolve this problem, one suggestion is to include the baseline risk in such communications. Four experiments examined (1) whether the presentation format (frequency vs. percentage) of the baseline risk matters for people's understanding of relative risk reductions and increases, and (2) whether this effect is different for low and high numerates. The experiments showed that relative risk reductions (increases) were often misunderstood as absolute risk reductions (increases), when the baseline risk was presented in a percentage format. For relative risk reductions, this misunderstanding led to an overestimation of actual treatment benefits; for relative risk increases, it led to an overestimation of treatment harms. Moreover, many participants ignored the baseline risk entirely or made calculation errors. Communicating baseline risk through frequency information particularly helped low numerates to correctly understand relative risk reductions (increases); high numerates interpreted relative risk statements mostly correctly, independent of format. Nevertheless, even a frequency format led a substantial proportion of participants to misinterpret the risk. If relative risk reductions and increases are used for communicating treatment benefits and harms, baseline risk should thus be provided in a frequency format. However, the findings suggest rethinking the use of relative risks in practice.

Introduction

Mammography screening reduces breast cancer mortality by 20% in women aged 50 and older. Certain drugs that lower cholesterol levels promise to cut the risk of stroke by 48% in patients with risk factors. In 1995, the UK Committee on Safety for Medicine released a warning that the third generation of the oral contraceptive pill increased the risk of life-threatening blood clots in legs or lungs by 100%, compared with the second generation pill.

What do these seemingly impressive numbers actually mean? The mammography screening and the cholesterol-lowering drug example communicate the benefit of a treatment in terms of a relative risk reduction; the contraceptive pill example states the potential harm of a treatment as a relative risk increase. A *relative risk reduction* (RRR) is defined as the difference in event rates (risk) in the control and treatment group, normalized by the event rate in the control group (the baseline risk) (Table 1). For instance, the 20% relative risk reduction of breast cancer mortality when participating in mammography screening is derived from the finding that 5 in 1,000 women aged 50 and older will die from breast cancer without screening, as opposed to 4 in 1,000 when participating in screening (Gøtzsche & Nielsen, 2006). Out of 100 patients with risk factors for stroke, 2.8 experience a stroke without taking a certain statin, while 1.5 who do take the drug experience a stroke or similar events (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007). A *relative risk increase* (RRI) is defined analogously (Table 1). In the contraceptive pill example, the risk of thrombosis increased from 1 out of 7,000 women taking the second-generation pill to 2 out of 7,000 women taking the third-generation pill—a relative risk increase of 100% (Gigerenzer & Gray, 2011; Williams, Kelly, Carvalho, & Feely, 1998).

An alternative way to express the same information is an absolute risk reduction, or absolute risk increase: the arithmetic difference between the event in a control group (baseline risk) and a treatment group (Table 1). In the mammography screening example, the *absolute risk reduction* would be 1 in 1,000 or, expressed in terms of percentages, 0.1%

(0.5% − 0.4%). The advantage of this measure—in contrast to a relative risk reduction—is that it also considers the total number of people at risk. For instance, a relative risk reduction of 50% may refer to a decrease from 2 in 10 in the control group to 1 in 10 in the treatment group, but it could also refer to a reduction from 2 in 1,000 to 1 in 1,000. By contrast, the absolute risk reduction reflects the difference between the two scenarios, being 10% in the former case, but 0.1% in the latter case.

Table 1

Relative and Absolute Risk Formats. The Example For the Risk Reductions Is Based on the Mammography Example (Mortality Reduction From 5 in 1,000 to 4 in 1,000 When Participating in Screening). The Example For the Risk Increase Measures Is Based on the "Pill Scare" (Increase of Thrombosis From 1 in 7,000 to 2 in 7,000 When Taking the Third-Generation Contraceptive Pill).

	Risk Reduction (RR)	Risk Increase (RI)
Relative (R)	$RRR = \frac{ER_{Control} - ER_{Treatment}}{ER_{Control}}$ Example: $\frac{0.5\% - 0.4\%}{0.5\%} = 20\%$	$RRI = \frac{ER_{Treatment} - ER_{Control}}{ER_{Control}}$ Example: $\frac{0.028\% - 0.014\%}{0.014\%} = 100\%$
Absolute (A)	$ARR = ER_{Control} - ER_{Treatment}$ Example: $0.5\% - 0.4\% = 0.1\%$	$ARI = ER_{Treatment} - ER_{Control}$ Example: $0.028\% - 0.014 = 0.014\%$

Note. $ER_{Control}$ = event rate in the control group (baseline risk); $ER_{Treatment}$ = event rate in the treatment group.

How relative vs. absolute risk reductions influence risk perception and decision making

Relative and absolute risks are based on the same statistical information (Table 1). However, the format in which treatment benefits and harms are communicated influences the evaluation of medical treatments. In general, treatment benefits presented as relative risk

reductions are more favorably evaluated than when presented as absolute risk reductions (e.g., Akl et al., 2011; Covey, 2007; Edwards Elwyn, Covey, Matthews, & Pill, 2001). For instance, in a survey regarding the willingness to participate in a fictitious cancer screening, 80% of participants were willing to participate when the treatment's benefits were presented in terms of a relative risk reduction. Only 53%, however, were willing to participate when presenting the screening's benefit in terms of an absolute risk reduction (Sarfati, Howden-Chapman, Woodward, & Salmond, 1998). When choosing between two equivalent treatments whose benefits were communicated in terms of either a relative or an absolute risk reduction, the majority of participants preferred the treatment whose benefits were presented as relative risk reduction (Malenka, Baron, Johansen, Wahrenberger, & Ross, 1993). Such findings are not limited to laypeople, but have also been observed with health professionals (Cranney & Walley, 1996; Forrow, Taylor, & Arnold, 1992; Mühlhauser, Kasper, Meyer, & Federation of European Nurses in Diabetes, 2006).

The negative consequences of using relative risk formats have been observed outside the lab as well. For instance, in 1995 empirical data from England and Wales showed a steep increase in the number of abortions, although this number had been on the decline since 1990. The rise was attributed to a decreased use of the contraceptive pill, resulting from the above-mentioned warning about the third-generation contraceptive pill (Gigerenzer & Gray, 2011; Williams et al., 1998).

Including baseline risk to improve understanding of relative risk reductions

Despite these findings, relative risk reductions remain the predominant format for communicating treatment benefits, be it in direct-to-consumer advertisements, brochures, and websites (Jorgensen & Gøtzsche, 2004; Moynihan, Bero, Ross-Degnan, Henry, Lee, et al., 2000; Slaytor & Ward, 1998) or within the medical community (Gigerenzer, Wegwarth, & Feufel, 2011; Schwartz, Woloshin, & Welch, 2006; Sedrakyan & Shih, 2007).

Why is that? First, health organizations like the Cochrane Collaboration still promote relative risk reductions, which may also explain why leading medical journals predominantly

provide them in their abstracts (Gigerenzer, Wegwarth, & Feufel, 2011; Sedrakyan & Shih, 2007). Although the use of absolute risk reductions is recommended in the "Summary of Findings" in the 2009 *Cochrane Handbook* (Higgins & Green, 2009), relative risk reductions "remain crucial because relative effects tend to be substantially more stable across risk groups than absolute risks" (Higgins & Green, 2009, p. 12-13). At the same time, omitting baseline risk is a major weakness; without this information one cannot make sense of the actual risk reduction.

This leads to the second argument for using relative risk reductions: Differences in the misperception of relative risk reductions are supposed to diminish when the baseline risk is also provided. Natter and Berry (2005) gave participants information about the risk reduction in a fictitious flu scenario, either in terms of relative or absolute risk reduction. Omitting baseline risk led to an overestimation of event rates regarding both the baseline risk in the population and the (reduced) risk in the treatment group. When people received information about the baseline risk, their estimates of treatment efficiency were more accurate. Schwartz, Woloshin, Black, and Welch (1997) asked women to estimate how many out of 1,000 women with and without mammography screening would die from breast cancer. Participants received the risk reduction either in relative or absolute terms. Including baseline risk improved accuracy for relative and absolute risk reductions alike. Yet at least two thirds of the participants were still incapable of giving the correct estimate, even when provided with baseline risk. In particular, participants with low numeracy abilities had major difficulties in solving the task.

Baseline risk: When and whom does it help?

The objective of our paper is to better understand how the inclusion of baseline risk helps people to interpret relative risk reductions correctly. Both Natter and Berry's (2005) and Schwartz and colleagues' (2007) studies focused on the effects of providing versus not

providing baseline risks. We extend prior research by identifying determinants and limitations for correctly interpreting relative risk reductions (or relative risk increases) when baseline risk is included. More specifically, we focus on two questions: (1) When including baseline risk to communicate relative risk information, does the presentation format of the baseline risk matter for understanding its meaning? (2) Does low and high numerates' understanding of relative risk information differ according to the presentation format of the baseline risk?

Risk communication and the role of the presentation format

The mathematical definition of a relative risk reduction is unambiguous. But is the perception and interpretation of it just as clear-cut? For instance, one explanation for the so-called "conjunction fallacy" (Tversky & Kahneman, 1983) is that the mathematical and logical definition of concepts like "probability" and (the logical) "and" have been interpreted differently in everyday contexts (Hertwig & Gigerenzer, 1999). A similar argument can be made with respect to relative risk reductions. For example, in 2009 about 35% of men aged 35 and older with diabetes in the US had heart disease or a stroke (CDC, 2011). Imagine a drug that reduces the risk of stroke in people with diabetes by 25% (Collins et al., 2004). What percentage of diabetics who take the drug will have heart disease?[1] The difficulty in making this inference lies in the potential ambiguity of the term "reduced by." What is meant is a relative risk reduction, implying that the event rate is reduced from 35% to 26.25%. However, an alternative (erroneous) interpretation is to consider the 25% reduction as referring to an absolute risk reduction. In that case, the risk reduction would refer to a decrease in 25 *percentage points*, meaning that the risk for heart disease is reduced from 35% to 10% when taking the drug. Thus, interpreting a relative risk reduction as an absolute risk reduction would lead to a gross overestimation of the treatment's effectiveness (26% vs. 10%).

[1] For illustrative purposes, we here assume that of the 35% of diabetics who had a stroke, nobody had received the drug. Furthermore, we assume that the drug is assumed to be equally effective for different risk factors.

The same argument can be made for interpretations of relative risk increase. Imagine that a certain treatment increases the risk of an undesirable side-effect by 25%, relative to a 35% baseline risk in a control group. The intended interpretation would imply an event rate of 43.75% in the treatment group. By contrast, erroneously interpreting the information as an absolute risk increase (i.e., an increase in percentage points) would yield an event rate of 60% in the treatment group—a strong overestimation of the actual risk.

While in the above example the baseline risk of stroke in diabetes patients without the drug was presented in percentage format (35% of the patients), one could also present the same information in frequency format: 350 out of 1,000 patients with diabetes experienced a stroke. Previous research found that people's capacity to reason with quantitative information depends on the external presentation format, for example whether information is provided in terms of probabilities or natural frequencies (Brase, 2008; Cosmides & Tooby, 1996; Galesic, Garcia-Retamero, & Gigerenzer, 2009; Gigerenzer & Hoffrage, 1995). When interpretating relative risk reductions with baseline risk, people's reasoning processes might be affected by the presentation format of the baseline risk. We hypothesize that one means of reducing the ambiguity in relative risk statements is to present the baseline risk in terms of frequencies (i.e., 350 out of 1,000). First, people often have difficulties in performing mathematical operations with percentages (Moser, 2002; Schwartz & Woloshin, 2000). Second, when baseline risk is presented as a frequency, it is not possible to directly subtract the risk reduction from the baseline value, as one must either convert the risk reduction into frequencies or the baseline risk into percentages. This may increase comprehension of the information by resolving potential ambiguity, either by clarifying the relevant reference class or by simplifying the computation (e.g., if the risk reduction is 20%, it may be easier to compute 20% of 300 than 20% of 30%).

The role of numeracy

Individual differences in the ability to comprehend and use numerical information affect how people understand relative risk reductions (Peters, 2008). The terms *statistical literacy* (Gigerenzer et al., 2007) and *numeracy* (Reyna, Nelson, Han, & Dieckmann., 2009) have been coined to define basic requirements for understanding numerical information. In general, the public—but also health professionals—are often described as being statistically illiterate, lacking the requisite skills for appropriately understanding statistical information (Gigerenzer et al., 2007; Lipkus, Samsa, & Rimer, 2001; Reyna et al., 2009). Numeracy is also an important moderator in risk perception and decision making, with high numerates being more precise in their numerical interpretation and less prone to framing effects (Garcia-Retamero & Galesic, 2010; Peters Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006). In line with the findings from Schwartz and colleagues (1997; see also Peters, 2008) regarding numeracy and the interpretation of relative risk reductions, we hypothesize that low numerates in particular might confuse relative risk reductions with absolute risk reductions.

Note that terms like *innumeracy* do not necessarily suggest a hard-wired incapacity to reason with numerical information. Rather, this capacity also depends on the presentation format that is used. For example, it has been argued that people perform better in solving Bayesian tasks when information is presented in natural frequencies compared with conditional probabilities (Cosmides & Tooby, 1995; Gigerenzer & Hoffrage, 1995), because natural frequencies more closely correspond to their experiences in the natural world (in which non-normalized frequency information is the "raw data"). Thus, finding ways to communicate information in a way that fosters people's understanding of statistical information is an important factor in risk communication.

Experiment 1a

The objective of Experiment 1 was to examine the conditions under which people correctly understand the meaning of a relative risk reduction (RRR interpretation) when the baseline risk is also provided. Specifically, our goal was to find out whether and when a relative risk reduction may be erroneously interpreted as an absolute risk reduction (ARR interpretation). Participants were presented with a fictitious medical scenario and a risk reduction statement similar to those commonly used in communicating medical information (e.g., "The drug reduces the risk by 20%"), together with information on the baseline risk. We hypothesized that such a risk statement is more likely to be erroneously interpreted as an absolute risk reduction—rather than a relative risk reduction—(i) when the baseline risk is presented in a percentage format and (ii) by people with low numeracy skills. By contrast, presenting baseline risk in terms of frequencies should increase the proportion of both low and high numerates who interpret the statement as referring to a relative risk decrease. A final goal of Experiment 1 was to identify participants' reasoning processes by asking them to describe how they arrived at their estimate.

Method

Participants. 101 participants (59.4% female; M_{age}=35, *SD*=11.5) were recruited through Amazon's MTurk for an online study; remuneration was $0.75. Participants were randomly assigned to one of two baseline risk conditions {percentage vs. frequency}.

Materials and Procedure. All participants were presented with the following hypothetical scenario:

> A study tested a new drug for diabetes. The study's aim was to find out whether and to what extent the new drug reduced the risk of heart disease. To evaluate the new drug, 2,000 patients with diabetes were tested. The patients were randomly assigned to two groups: 1,000 patients received the new drug and 1,000 patients received no drug. The patients receiving no drug served as a control group. After five years, the number of patients with heart diseases in each of the two groups was compared.

Subjects in the *baseline percentage condition* received then the following information:

> In the control group without the drug 30% of the patients had heart disease.
> The study showed that the new drug reduced the risk of heart disease by 20%.
> Please estimate how many patients in the group with the drug suffered from heart disease:
>
> O 24% out of 1,000 patients who received the drug had heart disease
> O 10% out of 1,000 patients who received the drug had heart disease

The first answer (24%) is the correct one, referring to a relative risk reduction of 20%. The second answer (10%) corresponds to an erroneous interpretation of the relative risk reduction as an absolute risk reduction.

Subjects in the *baseline frequency condition* received the same information, but here the baseline risk was presented in frequency format:

> In the control group without the drug, 300 out of 1,000 of the patients had heart disease.
> The study showed that the new drug reduced the risk of heart disease by 20%.
> Please estimate how many patients in the group with the drug suffered from heart disease:
>
> O 240 out of 1,000 patients who received the drug had heart disease
> O 100 out of 1,000 patients who received the drug had heart disease

Note that all participants received the same risk reduction statement ("The study showed that the new drug reduced the risk of heart disease by 20%"). The crucial difference was whether baseline risk was presented as percentage or frequency. The order of the two answers was randomized. After participants selected their answer and provided a written explanation of their decision, their numeracy was assessed using the 11-item numeracy scale from Lipkus and colleagues (2001), with the items presented in random order.

Results and Discussion

The results show a strong influence of presentation format on the interpretation of the risk reduction statement (Figure 1). In the percentage condition, only 51% (24/47) of participants correctly interpreted the statement as a relative risk reduction, estimating that

24% out of 1,000 patients who did receive the drug would suffer from heart disease. The other half of the subjects (49%) estimated that the drug would reduce the risk of heart disease to 10%. Thus, these participants erroneously understood the statement as an absolute risk reduction, thereby strongly overestimating the drug's effectiveness.

In stark contrast, 83% (45/51) of participants in the frequency condition assumed that the drug would reduce the event rate from 300 to 240, thereby correctly interpreting the statement as a relative risk reduction (two-tail binomial, p=.001). Thus, the frequency format helped people to make sound inferences regarding what the risk reduction statement implied.

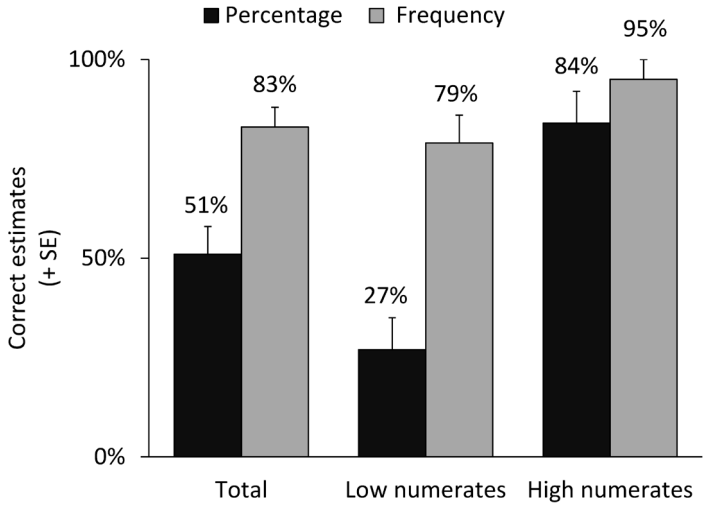


Figure 1. Results of Experiment 1a. In the percentage condition, the event rate in the control group (baseline risk) was set to 30%, in the frequency condition to 300 out of 1,000. "Total" includes all subjects' interpretations in the percentage and frequency condition (n=101). Low numerates are those with a numeracy score

Influence of numeracy. Does people's numeracy affect their interpretation of relative risk information? We conducted a median split to categorize participants as low numerates (≤ 9 items correct) and high numerates (≥10 items correct) (cf. Peters et al., 2006; Galesic et al., 2009). For low numerates, a strong effect of presentation format was observed (Figure 1). In the percentage condition, only 27% (7/26) of low numeracy participants correctly interpreted

the information as referring to a relative risk reduction (two-tail binomial, $p = .03$). By contrast, when baseline risk was presented in terms of frequencies, 79% (26/33) correctly interpreted the statement (two-tail binomial, $p=.001$). The inferences of the high numerates were only weakly affected by presentation format. In both conditions, most participants correctly interpreted the statement as a relative risk reduction: 84% (16/19; two-tail binomial, $p = .004$) in the percentage condition and 95% (19/20; two-tail binomial, $p = .0004$) in the frequency condition.

The finding that only high numerates were likely to understand the health information in the intended way suggests that numeracy plays an important role when baseline risk is presented in terms of percentages. By contrast, numeracy was largely irrelevant when baseline risk was presented in terms of frequencies; in this case low numerates also tended to correctly understand the statement as a relative risk reduction.

Reasoning analysis. The final analysis concerns participants' written descriptions of their reasoning processes. Answers were coded according to four categories: (1) Did the participant explicitly refer to a numerical calculation to justify the selected answer ("calculation")? For example, a participant selecting an RRR interpretation might explicitly state that 24% = 30% - (30% × 0.2) or 30% × 0.8. Conversely, a participant making an ARR interpretation might report 10% = 30% - 20%. (2) Did a participant use a mental short cut to justify the selected answer? For example, a participant stated: "I am bad at math; however, 10% seems way too low of a number compared to 24%." Hence, one might compare the two answer options and select the one that intuitively appears more plausible. (3) Did the participant guess? Some subjects explicitly indicated that they simply guessed which answer might be correct. (4) A fourth category subsumes participants whose reasoning was not identifiable ("not identified").

Table 2 shows the results of this analysis. Most participants who made an RRR interpretation explicitly noted the corresponding calculation (78% in the percentage and 73% in the frequency condition). Whereas only 9% in the percentage condition used a short cut, 22% in the frequency condition did so. Among participants who used a short cut in the frequency condition, most interpreted the 20% risk reduction as the event rate of 200 out of 1,000 in the treatment group and then selected the RRR interpretation, which was closer to 200 than the ARR interpretation. (This interpretation will be addressed in more detail in Experiment 2.) With respect to the ARR interpretations, 64% provided the formula to arrive at their answer, 9% used a short cut, and 18% guessed when the baseline risk was given as a percentage. Only eight participants made an ARR interpretation in the frequency condition, making it difficult to interpret the data.

As the analysis shows, most people actually used a formal reasoning process to arrive at their estimate. This was the case regardless of whether they interpreted the statement as referring to a relative or absolute risk reduction.

Table 2

Classification of Participants' Written Descriptions of Their Reasoning Processes in Experiment 1a.

	Baseline risk as percentage				Baseline risk as frequency			
Interpretation	Calculation	Short Cut	Guess	Unidentified	Calculcation	Short Cut	Guess	Unidentified
RRR	78%	9%	4%	9%	73%	22%	4%	–
(*n*=68)	(18/23)	(2/23)	(1/23)	(2/23)	(33/45)	(10/45)	(2/45)	
ARR	64%	9 %	18%	9%	38%	25%	0%	38%
(*n*=30)	(14/22)	(2/22)	(4/22)	(2/22)	(3/8)	(2/8)	(0/0)	(3/8)

Note. RRR = relative risk reduction, ARR = absolute risk reduction. Three participants did not provide a justification of their answer and were excluded from this analysis.

Experiment 1b

The goal of Experiment 1b was to examine how participants interpret a statement about a relative risk *increase* (RRI interpretation) (Table 1). Again, such a statement might be falsely interpreted as an absolute risk increase (ARI interpretation), that is, an increase in percentage points. We hypothesized similar results to those in Experiment 1a, namely an influence of presentation format and low numerates being more likely to interpret the relative risk increase as an absolute increase, particularly when the baseline risk is presented as a percentage.

Method

Participants. Seventy-seven participants (58.4 % female; M_{age}=33.7 years, *SD*=12.4) recruited via Amazon MTurk took part in an online study. Participants were randomly assigned to one of two baseline risk conditions {percentage vs. frequency}; they were paid $0.75 for participation.

Materials and Procedure. The experimental procedure was virtually identical to Experiment 1a. However, participants were presented with a risk increase rather than risk reduction scenario. Specifically, participants were told that a drug for diabetes patients would *increase* the risk of heart disease by 20%. The baseline risk in the control group (i.e., patients who do not take the drug) was set to 30% out of 1,000 (*baseline percentage condition*) or 300 out of 1,000 (*baseline frequency condition*). Participants had to choose between two possible answers: a relative risk increase interpretation (36% in the percentage condition, 360 out of 1,000 in the frequency condition) and an absolute risk increase interpretation (50% and 500 out of 1,000, respectively). The order of answers was randomized.

Results and Discussion

Figure 2 shows that a number of participants erroneously interpreted the statement as an absolute risk increase, but that there was relatively little effect of presentation format. In the percentage condition, 67% (26/39) of participants gave an answer consistent with a

relative risk increase (two-tail binomial, $p = .05$). In the frequency condition, 76% (29/38) interpreted the statement as a relative risk increase (two-tail binomial, $p = .002$). Thus, regardless of the presentation format a majority of participants interpreted the statement as referring to a relative risk increase. One explanation for the weak influence of presentation format finding is that subjects may have relied on their background knowledge, concluding that a risk increase to 50% for a severe disease it too high.

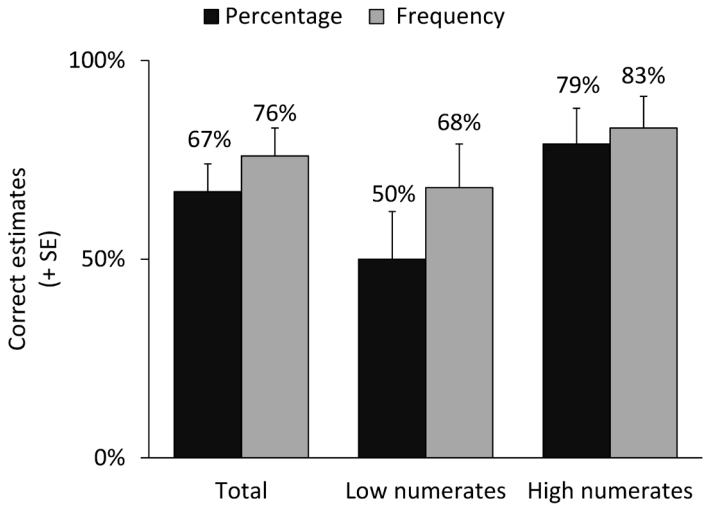


Figure 2. Results of Experiment 1b. In the percentage condition, the event rate in the control group was set to 30%, in the frequency condition to 300 out of 1,000. "Total" includes all subjects' interpretations in the percentage and frequency condition (n=77). Low numerates are those with a numeracy score

Numeracy analysis. As in Exp. 1a, we conducted a median split to categorize participants as low or high numerates. When baseline risk was conveyed as percentage, 50% (6/12) of low numerates correctly interpreted the statement (Figure 2). The lack of an overall effect of presentation format, as was seen in Exp. 1a for a relative risk reduction, is thus due to a higher number of low numerates correctly interpreting the risk increase statement, even when baseline risk was conveyed as percentage. At the same time, frequency information increased the proportion of low numerates' correct interpretations to 68%, although responses were at chance level (13/19; two-tail binomial, $p = .17$). As in Exp. 1a, presentation format

did not matter for the high numerates. When baseline risk was given as percentage, 79% (19/24; two-tail binomial, $p = .007$) gave the correct answer; in the frequency condition 83% (15/18; two-tail binomial, $p = .008$) correctly opted for the relative risk increase. Thus, similar to Experiment 1a, low numerates benefited from presenting the event rate in frequency format, while high numerates correctly understood the statement as a relative risk increase, independent of presentation format.

Reasoning analysis. Table 2 shows the results of the reasoning analysis. In the percentage conditions, 65% provided the respective calculation when following a RRI interpretation and 73% when following an ARI interpretation (Table 3). For the frequency condition, the respective proportions were 79% and 67%. A short cut was used in 15% in the percentage and 17% in frequency condition in the RRI interpretations. Many of those who used a short cut referred to the fact that, given the base line risk, a 20% risk increase to 50% (or 500) out of 1,000 would be too high.

Table 3

Classification of Participants' Written Descriptions of Their Reasoning Processes in Experiment 1b.

	Baseline risk as percentage				Baseline risk as frequency			
Interpretation	Calculation	Short Cut	Guess	Unidentified	Calculation	Short Cut	Guess	Unidentified
RRI (*n*=55)	65% (17/26)	15% (4/26)	8% (2/26)	12% (3/26)	79% (23/29)	17% (5/29)	3% (1/29)	–
ARI (*n*=20)	73% (8/11)	9% (1/11)	18% (2/11)	–	67% (6/9)	11% (1/9)	22% (2/9)	–

Note. RRI=relative risk increase, ARI=absolute risk increase. Two subjects did not provide a reason and were excluded from this analysis.

Experiment 2a

Experiment 1a showed that people tend to interpret a relative risk reduction as an absolute risk reduction when (i) the event rate in the control is presented in percentage format, and (ii) people have low numeracy skills. A similar, albeit somewhat weaker pattern was observed with risk increase statements. In both experiments, the answer options were restricted to a RRR (RRI) and ARR (ARI) interpretation only. We designed Experiments 2a,b to further explore the range of possible interpretations and reasoning processes. For instance, what both previously tested interpretations have in common is that estimates are based on a comparison between treatment and control group. Although we explicitly stated that the evaluation of the drug was based on such a comparison, its intuitive consideration requires a basic understanding of the idea of randomized controlled trials.

Are laypeople familiar with this concept and do they interpret the risk reduction statement accordingly? For example, one might ignore the information about the event rate in the control group and assume instead that the 20% risk reduction refers to the event rate in the treatment group (ER interpretation). In this case, the statement would not be interpreted as a reduction relative to the event rate in the control group (baseline risk), but as directly denoting the event rate in the treatment group. This interpretation is suggested by the written statements of some participants in Experiment 1a, who used a short cut based on such an interpretation (e.g., “20 percent of 1,000 is 200, so 240 is closer to 20 percent than the other 100”). Since we did not offer an ER interpretation, they may have opted for a relative risk interpretation, which was numerically closer to their intended interpretation.

We consider these three interpretations (RRR, ARR and ER interpretation) as *conceptual* interpretations of a relative risk reduction statement, since they refer to three qualitatively different understandings. These interpretations must be distinguished from judgments resulting from mere computational errors. For instance, many people have difficulties converting percentages into frequencies and vice versa (Schwartz & Woloshin,

2009). In our example, 20% out of 1,000 corresponds to 200 in 1,000. However, a false conversion could result in the answer 20 out of 1,000 (which would be equivalent to a 2% risk reduction). Along with an ARR interpretation, this may then lead participants to interpret the 20% risk reduction as 300 – 20 = 280 out of 1,000 in the frequency condition, or 30% − 2% = 28% out of 1,000 in the percentage condition (Error I). Likewise, people may interpret the risk reduction statement as referring to the event rate in the treatment group, but conduct an error by estimating the event rate in the treatment group as 20 out of 1,000 or 2% out of 1,000 (Error II). Figure 3 outlines the different interpretations.

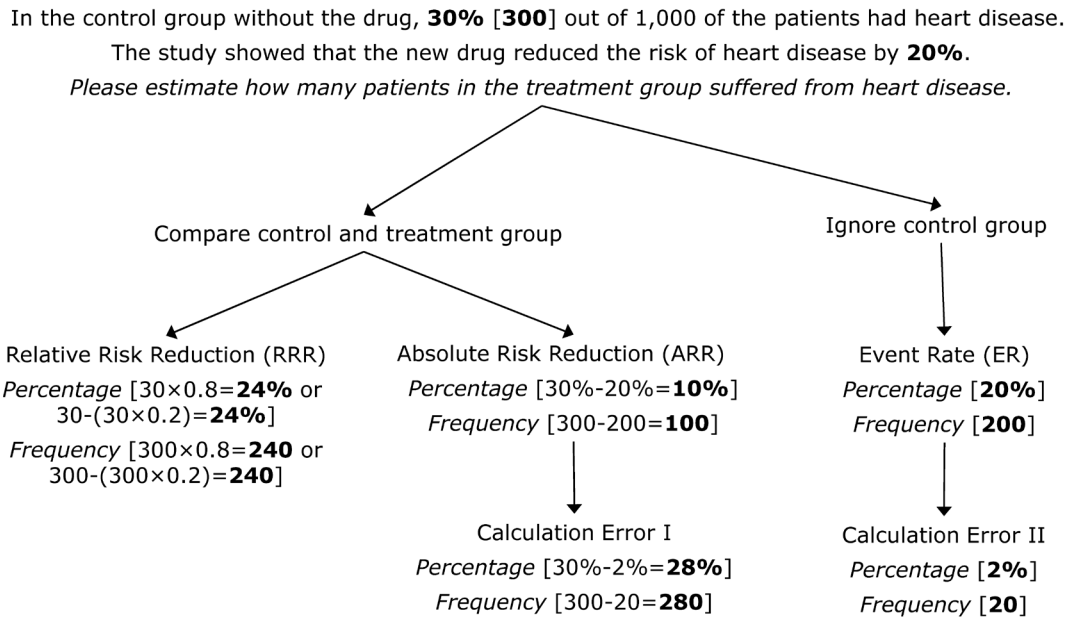


Figure 3. Five potential interpretations of a relative risk reduction statement, and estimates resulting from different interpretations. Depending on whether the event rate in the control group is considered or not, three different conceptual interpretations result (RRR interpretation, ARR interpretation, ER interpretation). In addition, two possible errors (Error I, Error II) are illustrated.

A second goal of the current study was to examine how participants interpret the relative risk reduction when it is numerically larger than the baseline risk (for instance, when the baseline risk is 30% and the risk reduction is 40%). Such a scenario effectively rules out

the possibility of interpreting the relative risk reduction as absolute difference: simply subtracting the two numbers would result in a negative value. As a consequence, the people in question should revise their initial interpretation and reconsider what the risk statement implies.

Method

Participants. 221 participants (53% female, M_{age}=35 years, *SD*=11.8) were recruited through Amazon's MTurk for participating in an online study; they were paid $0.75. Participants were randomly assigned to one of four conditions: one of two presentation formats for baseline risk {percentage vs. frequency} × one of two levels of risk reduction {low=20% vs. high=40%}.

Materials and Procedure. We used the same scenario as in Experiment 1a, according to which a new drug for diabetes patients reduces the risk of heart disease by 20% (see Appendix for full instructions), with two modifications. First, we provided participants with five answer options (see Figure 3 and Appendix). For the ER interpretation the answer was 20% out of 1,000 (percentage condition) and 200 out of 1,000 (frequency condition). The answers corresponding to Error I were 28% out of 1,000 in the percentage condition and 280 in 1,000 in the frequency condition. Error II corresponded to 2% out of 1,000 in the percentage condition and 20 in 1,000 in the frequency condition. All answers were presented in random order.

Second, we varied the size of the risk reduction, being either 20% or 40%. Since the event rate in the control group was fixed at 30% [300] out of 1,000, the high risk reduction level precludes an ARR interpretation, which would yield a negative event rate in the treatment group (e.g., 30% − 40% = −10%). The question of interest in this case was whether the majority of participants would opt for the relative reduction interpretation or switch to one of the remaining answers, such as favoring the ER interpretation or Error I, which

conceptually resembles an ARR interpretation. The five answers for the 40% condition were adjusted for this risk reduction. As in the previous studies, the baseline risk in the control group was presented either as percentage (30% out of 1,000) or as frequency (300 out of 1,000).

Results and Discussion

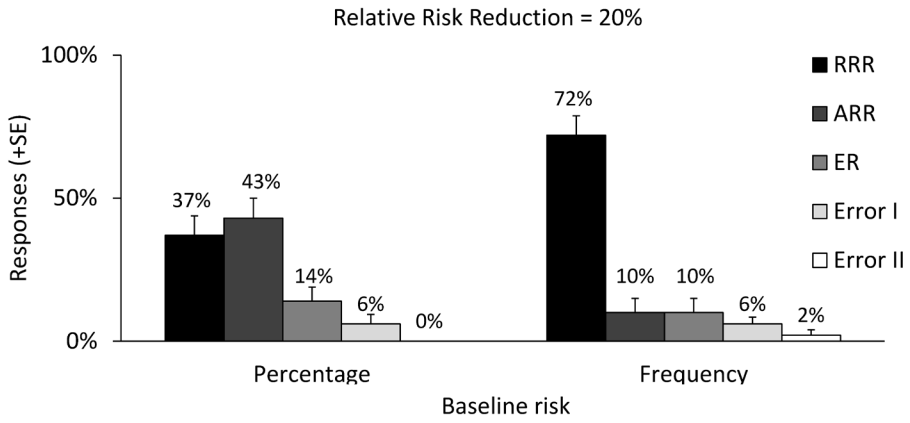
Figure 4 shows the results of Experiment 2a. When the risk reduction was 20% (Figure 4a), a strong influence of presentation format was observed. When baseline risk was communicated as percentage, the two most frequent answers were an ARR interpretation (43%) and an RRR interpretation (37%). By contrast, when baseline risk was presented as frequency, 72% of subjects correctly interpreted the statement as a relative risk reduction; only 10% chose an ARR interpretation. This corroborates the findings of Experiment 1a, showing that using frequencies to provide baseline risk helps people to understand the intended meaning of a risk reduction statement.

What about the alternative interpretations? In both conditions, a similar percentage of participants interpreted the risk reduction statement as directly referring to the event rate in the treatment group (ER interpretation): 14% in the percentage condition and 10% participants in the frequency condition. Few errors were observed in either condition (Figure 4a).

Does ruling out a meaningful ARR interpretation change the interpretation pattern? Figure 4b shows the results for the conditions in which the baseline risk was 30% [300] out of 1,000 and the risk reduction was 40%. Overall, no substantial difference between the two presentation formats was observed. Regardless of whether the event rate in the control group was presented as percentage or frequency, the (correct) RRR interpretation was the most common answer (60% in the percentage condition and 66% in the frequency condition). Very

few people (4 out of 123 across both conditions) gave an ARR interpretation, meaning that they judged the risk in the treatment group to be zero.

a)



b)

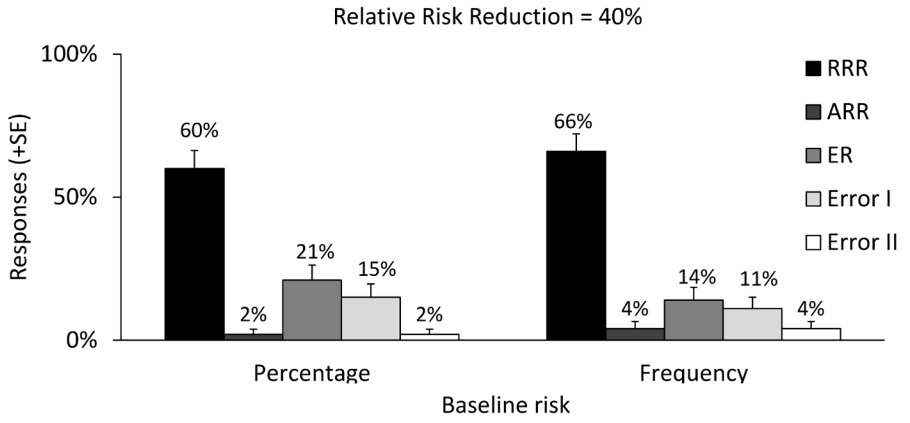


Figure 4. Results of Experiment 2a (*N*=221). (a) Interpretations in the 20% risk reduction condition (*n=98*). (b) Interpretations in the 40% risk reduction condition (*n=123*). In all conditions the baseline risk was set to 30% (percentage condition) or 300 (frequency condition) out of 1,000.

While the high risk reduction statement eliminated the ARR interpretations as expected, the interesting finding is that in both conditions an increased proportion of people opted for the ER interpretation (21% in the percentage condition and 14% of participants in the frequency condition). These participants considered the risk reduction as directly referring to the event rate in the treatment group. In addition, more participants made computational

errors, accounting for 17% of responses in the percentage condition and 15% in the frequency condition (aggregating across the two types of errors, with more people making Error I, see Figure 4b).

Overall, eliminating an ARR interpretation as meaningful response strongly increased the number of correct interpretations even when baseline risk was presented as percentage, but we also observed that a considerable number of people chose the event rate interpretation or performed computational errors when assessing the treatment effect.

Numeracy. In all four conditions, strong differences between low and high numerates were observed (Figure 5). Most of the high numerates correctly understood the relative risk reduction statement, except for the condition in which the baseline risk was 30% and the reduction was 20%, in which only 53% of high numeracy participants gave the correct answer. In the low risk reduction conditions, the most common misunderstanding was the ARR interpretation, with 33% in the percentage condition and 12% in the frequency condition. In the low risk reduction percentage condition, only 11% of the low numerates followed an RRR interpretation, while 61% interpreted it as an absolute risk reduction. Between 44% and 47% followed an RRR interpretation in the other three conditions. The second most common answer was the ER interpretation, made by about one quarter (22% - 28%) of the low numerates. Error I was particularly frequent in the 40% risk reduction conditions, (20% in the percentage and 17% in the frequency condition).

In sum, low numerates had great difficulties interpreting the risk reduction statement. This was largely independent of the presentation format and the magnitude of the risk reduction. With the exception of the low risk percentage condition, high numerates performed equally well across conditions. Eliminating a meaningful ARR interpretation led to a higher proportion of answering corresponding to an ER interpretation or Error I.

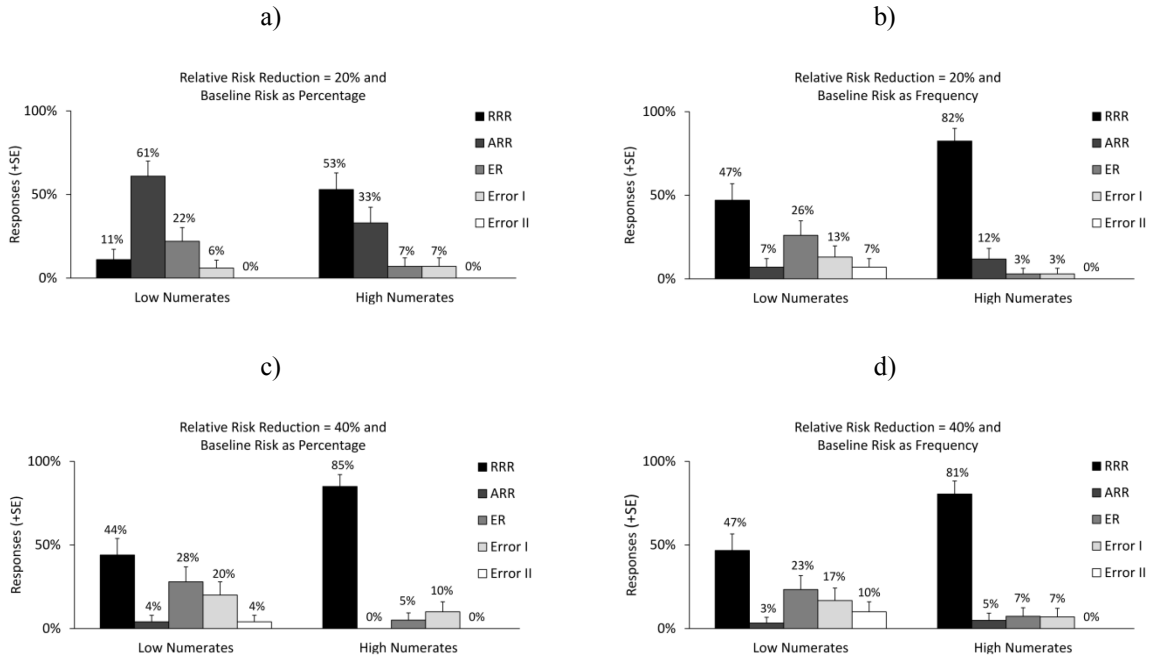


Figure 5. Results of Experiment 2a for low and high numerates (*N*=210). (a) Interpretations in the 20% percentage condition (*n=45*), (b) interpretations in the 20% frequency condition (*n=49*), (c) interpretations in the 40% percentage condition (*n=45*), (d) Judgments in the 40% frequency condition (*n=71*).

Reasoning analysis. As in the previous studies, we analyzed participants' written explanations of their estimates and reasoning processes (Table 4). Having found no differences in the reasoning process between the low and high risk conditions, we summarized the results and presented them according to presentation format of baseline risk (percentage vs. frequency). At least two thirds of the participants following either one of the conceptual interpretations (RRR, ARR, or ER interpretation) explicitly stated the respective calculation. This was independent of the presentation format. People making an RRR interpretation used short cuts more frequently (16% in the percentage and 13% in the frequency format), whereas those making an ER interpretation guessed more often, particularly in the percentage condition (27%). A common short cut of the RRR interpreters in the high risk reduction condition was to roughly consider a 50% risk reduction and "add a little" to the treatment group's event rate in order to arrive at a 40% risk reduction. When it

comes to the errors—which were in general rare in all four conditions—the pattern becomes slightly unclear. Few participants referred to the actual formula, but instead used a short cut or guessed. Hence, some participants who selected an answer option based on the error might have had another reasoning process, as we originally hypothesized.

Table 4

Classification of Participants' Written Descriptions of Their Reasoning Processes in Experiment 2a, Averaged Across the 20% and 40% Risk Reduction Condition.

	Baseline risk as percentage				Baseline risk as frequency			
Interpretation	Calculation	Short Cut	Guess	Unidentified	Calculation	Short Cut	Guess	Unidentified
RRR (*n*=127)	73% (33/45)	16% (7/45)	9% (4/45)	2.2% (1/45)	81% (66/82)	13% (11/82)	5% (4/82)	1% (1/82)
ARR (*n*=29)	76% (16/21)	10% (2/21)	10% (2/21)	4.8% (1/21)	75% (6/8)	25% (2/8)	–	–
ER (*n*=30)	67% (10/15)	7% (1/15)	27% (4/15)	–	67% (10/15)	7% (1/15)	13% (2/15)	13% (2/15)
Error I (*n*=20)	33% (3/9)	11% (1/9)	56% (5/9)	–	27% (3/11)	36% (4/11)	27% (3/11)	9% (1/11)
Error II (*n*=5)	–	100% (1/1)	–	–	50% (2/4)	25% (1/4)	25% (1/4)	–

Note. RRR = relative risk reduction, ARR = absolute risk reduction, ER = event rate interpretation. Ten subjects did not provide a reason and were excluded from this analysis.

Experiment 2b

In Experiment 2b, we examined participants' interpretations of a relative risk increase, with multiple answer options and under conditions in which interpreting the statement as an absolute increase was not meaningful. We used and adapted the five answer options from Experiment 2a and manipulated the size of the risk increase (20% vs. 80%). Because the sum of the baseline risk (30% and 300 out of 1,000) exceeds 100% or 1,000 out

of 1,000, respectively, high risk increase of 80% should reduce the likelihood of an absolute risk increase interpretation.

Method

Participants. 180 participants (61% female, M_{age}=29 years, *SD*=8.9) were recruited through Amazon's MTurk for an online study; they were paid $0.75. Participants were randomly assigned to one of four conditions: one of two baseline risk presentation formats {Percentage vs. Frequency} × one of two levels of risk increase {low=20% vs. high=80%}.

Materials and Procedure. We used the same medical scenario as in Experiment 1b, in which a drug for diabetes patients had been shown to increase the risk of heart disease. The only difference was that participants had five answer options available (see Appendix).

Results and Discussion

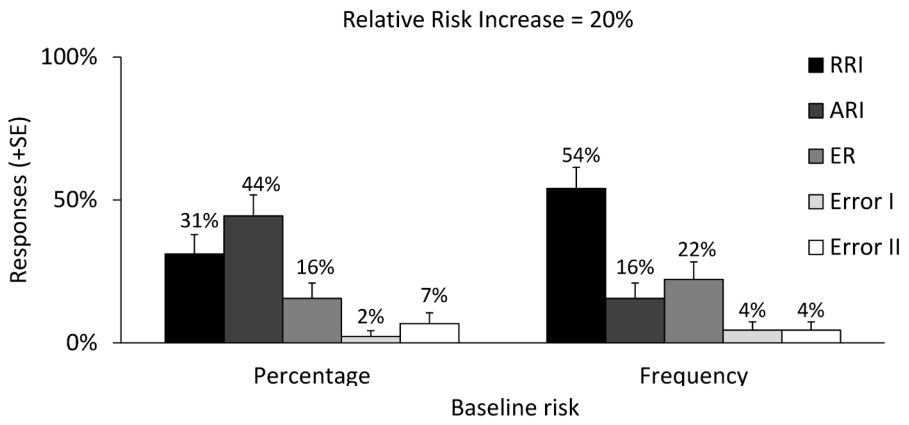
Overall, a lower proportion of correct responses was observed than in the risk reduction scenario. In the low risk increase condition, when baseline risk was given as a percentage, only 31% correctly interpreted the statement as a relative risk increase (Figure 6a); 44% of participants interpreted it as an absolute risk increase, and 16% interpreted it as referring to the event rate in the treatment group (ER interpretation).

When the baseline risk was conveyed as a frequency, the most common answer (54%) corresponded to a relative risk increase; about 16% considered the statement as referring to an absolute increase. Interestingly, the second most common answer (22%) was based on assuming the statement to directly refer to the event rate in the treatment group (ER interpretation). Errors were relatively rare, regardless of presentation format (Figure 6a).

What happened in the high risk increase condition? In this case, regardless of whether baseline risk was given in terms of percentages or frequencies, the most common answer corresponded—correctly—to a relative increase. However, in terms of absolute numbers, only about half of the subjects correctly understood the risk information. Excluding the absolute

risk increase as meaningful estimate led to a substantial proportion of people either assuming the risk statement to refer directly to the event rate in the treatment group or following Error I (Figure 6b). Notably, a higher proportion of errors was observed than in the low risk condition.

a)



b)

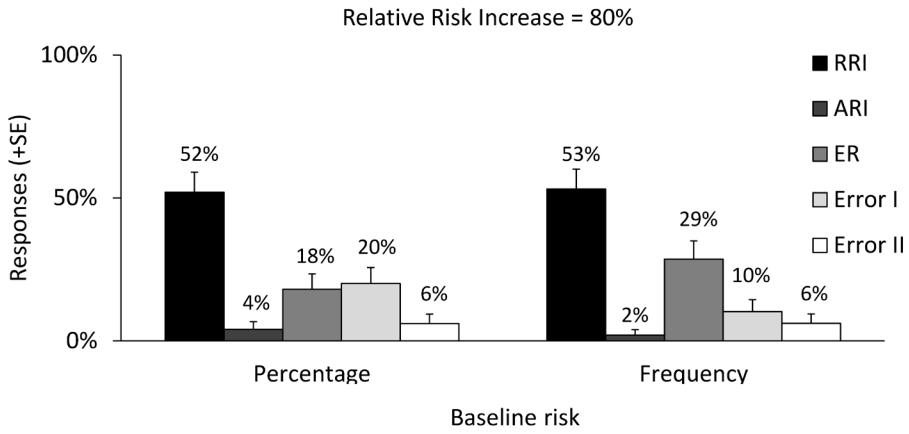


Figure 6. Results of Experiment 2b for (a) the 20% risk increase condition in the percentage and frequency condition (n=90) and for (b) the 80% risk increase condition (n=99).

Numeracy analysis. The numeracy analyses yielded clear differences between people with low and high numeracy skills (Figure 7). The key finding is that low numeracy people were particularly likely to misinterpret the risk statement as referring to an absolute increase. Only when this interpretation was not meaningful (i.e., in the conditions in which adding up baseline risk and increase would exceed 100%) did a substantial proportion of people

correctly interpret the statement as a relative increase. Interestingly, many low numerates interpreted the risk statement as denoting the actual risk level in the treatment group (ER interpretation), suggesting that they ignored the fact that the treatment's effectiveness was relative to a control group. Moreover, a high proportion of them were observed to follow Error I, primarily in the high risk percentage conditions.

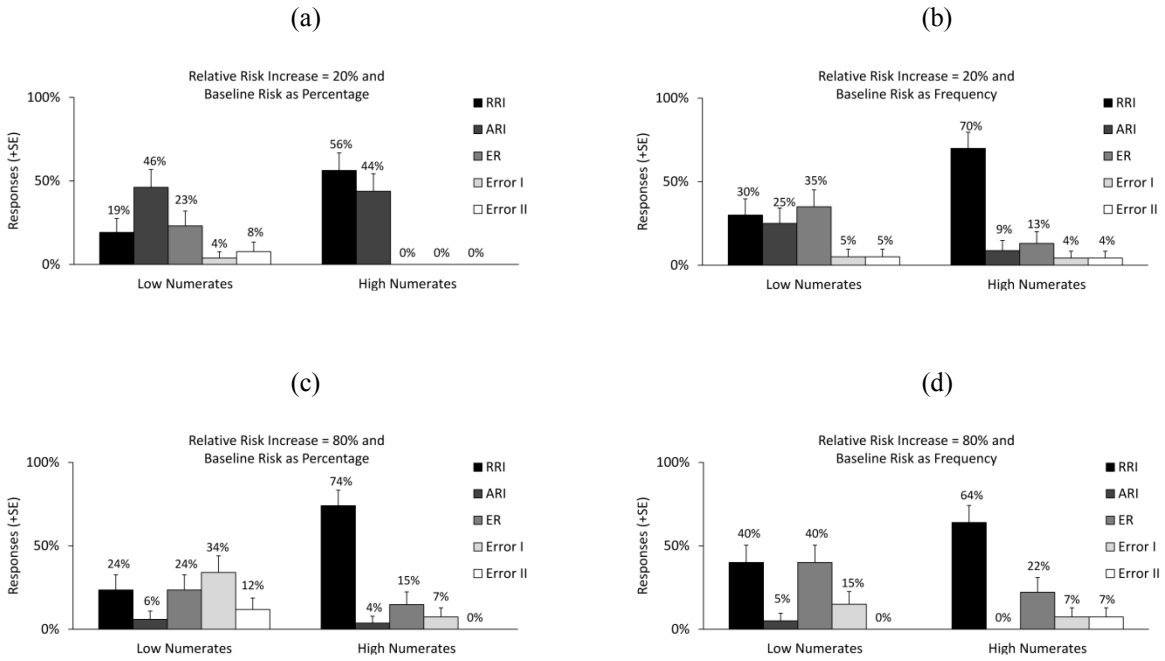


Figure 7. Results of Experiment 2b for low and high numerates (*N*=176). (a) Interpretations in the low risk (20%) percentage format condition (*n*=42), (b) interpretations in the 20% frequency condition (*n*=43), (c) interpretations in the 40% percentage condition (*n*=44), and (d) interpretations in the 40% frequency condition (*n*=47).

What about high numeracy people? High numerates had most difficulties in the low risk increase condition when baseline risk was provided as percentage, with 56% making a RRI interpretation and 44% making an ARI interpretation (Figure 7). Between 64% and 70% of high numerates followed the intended RRI interpretation in the other three conditions. For high numerates, the most common misinterpretation was the ER interpretation (between 13 and 22%,

depending on condition). Few errors were observed.

The observed pattern resembles the results of the previous experiments. Low numerates had difficulties interpreting the risk increase statement as intended. However, high numerates also showed a stronger tendency to misunderstand the risk increase statement than when given the risk reduction scenario.

Reasoning analysis. Table 5 summarizes results of participants' reasoning in the percentage and frequency condition, aggregated for the low and high risk conditions.

Table 5

Classification of Participants' Written Descriptions of Their Reasoning Processess in Experiment 2b, Averaged Across the 20% and 80% Risk Increase Condition.

	Baseline risk as percentage				Baseline risk as frequency			
Interpretation	Calculation	Short Cut	Guess	Unidentified	Calculation	Short Cut	Guess	Unidentified
RRI (*n*=87)	79% (30/38)	16% (6/38)	5% (2/38)	–	83% (41/49)	10 % (5/49)	0%	6% (3/49)
ARI (*n*=29)	81% (17/21)	5% (1/21)	5% (1/21)	10% (2/21)	63% (5/8)	25% (2/8)	13% (1/8)	–
ER (*n*=37)	79% (10/13)	23% (3/13)	0%	–	75% (18/24)	17% (4/24)	8% (2/24)	–
Error I (*n*=15)	50% (4/8)	25% (2/8)	25% (2/8)	–	43% (3/7)	14 % (1/7)	14% (1/7)	29% (2/7)
Error II (*n*=9)	100% (4/4)	–	–	–	20% (1/5)	80% (4/5)	–	–

Note. RRI = relative risk increase, ARI = absolute risk increase, ER = event rate interpretation. Eleven subjects did not provide a reason and were excluded from this analysis.

The majority of participants who made one of the three conceptual interpretations explicitly stated the formula of the respective interpretation. This pattern is similar in the

percentage and frequency conditions. The small sample size for the Error interpretations limits generalizations of this pattern.

General Discussion

Our studies show that the correct understanding of relative risk information with baseline risk depends (1) on the presentation format of the baseline risk and (2) people's numeracy skills. A large proportion of participants misunderstood relative risk information when the baseline risk was presented in a percentage format. Specifically, when the baseline risk was larger than the relative risk reduction, participants tended to interpret a relative risk reduction as an absolute reduction, thereby overestimating the actual treatment effect. Communicating baseline risk in a frequency format improved subjects' comprehension. Our findings also show an influence of people's numeracy skills: Whereas high numerates tended to interpret relative risk statements correctly, independent of format, low numerates were likely to misunderstand relative risk information. These participants particularly benefited from baseline risk being conveyed in a frequency format. Finally, even when the possibility of misinterpreting a relative risk reduction as an absolute reduction was ruled out because the reduction was larger than the baseline risk, other misunderstandings were observed. For instance, many participants understood the risk reduction as directly referring to the event rate in the treatment group. Results for relative risk increases were similar, but showed a slightly larger proportion of misunderstandings for high numerates as well.

Previous research has documented that treatments framed in terms of relative risk reductions are preferred over equivalent treatments framed in terms of absolute risk reductions (e.g., Akl et al., 2011; Covey, 2007; Edwards et al., 2001). Our analyses offer an explanation as to why this happens: Most misinterpretations of relative risk statements lead to a strong overestimation of the actual treatment effect. For instance, in Experiment 1a and 2a, incorrectly interpreting the relative reduction as an absolute reduction leads to an

overestimation of the treatment effect by 14 percentage points. Even when the baseline risk was presented in frequency format and an absolute risk reduction interpretation was ruled out, about one third of participants could not provide the correct estimate. Conceptual misunderstandings (e.g., ignoring the control group) as well as computational errors remain a source of misperception.

The problem of using relative risk reductions and increases in practice becomes even more evident when considering that information about baseline risk is usually omitted in health messages. Schwartz, Woloshin, Dvorin, and Welch (2006) found that leading medical journals often report ratio measures without explicitly stating the baseline risk, which is left unclear. This adds an additional source of ambiguity and uncertainty. For instance, Natter and Berry (2005) found that if no baseline risk is provided, people strongly overestimate risks.

In summary, seemingly simple mathematical concepts like relative risk reductions and increases may result in misinformed and misled patients, even when information on the baseline risk is provided. Hence, the use of relative risk reduction contributes to the phenomena of biased reporting in medical journals, biased reporting in pamphlets, and biased reporting the media—three out of the "seven sins in health care" (Gigerenzer & Gray, 2011). Even more disconcerting is that in one out of three abstracts published in leading medical journals—JAMA, BMJ, and The Lancet—treatment benefits are reported as relative risk reduction but potential harms are reported in absolute numbers, which makes them appear small in relation; a phenomenon called *mismatched framing* (Gigerenzer et al. 2010; Sedrakyan & Shih, 2007). Likewise, media analyses show that many pamphlets, websites, and newspapers tend to either communicate no numerical information at all about treatments or only in terms of relative risk reductions (Bodemer et al., 2012; Jorgensen & Gøtzsche, 2004; Moynihan et al., 2000; Slaytor & Ward, 1998). As a consequence, biased risk communication undermines the idea of shared decision making—patients' active participation in making

informed health decisions—in two ways. First, treatment recommendations by health professionals and policy-makers are based on nontransparent and incomplete information. Second, patients cannot correctly assess and evaluate treatment benefits and harms, and may select treatments that they would not favor had they been properly informed. The consequences are far-reaching and lead to poor decision making both on the individual and on a public health level, with negative effects for health and health care systems (Gigerenzer & Gray, 2011).

A problem related to biased reporting is the fact that large proportions of the public have difficulties understanding statistical information and lack the ability to identify nontransparent formats and translate them into more transparent statistics (Peters, 2008; Gigerenzer et al., 2007; Reyna et al., 2009). For instance, Galesic and Garcia-Retamero (2010) assessed numeracy on samples in the USA and Germany. The study not only identified large differences in numeracy skills between lower and higher educated people but also showed that numerical and statistical literacy is prevalent even in well-educated samples (Lipkus et al., 2001; Schwartz et al., 1997). These findings stress the importance of communicating health information as transparently and intuitively as possible. An alternative format for communicating benefits and harms is the *fact box* (Arkes & Gaissmaier, in press; Bodemer et al., 2012; Schwartz & Woloshin, 2009). Figure 8 shows an example of a fact box for breast cancer screening. It provides information on how many out of 1,000 women who do not participate in mammography screening and how many women who do participate in routine screening will die from breast cancer in the next ten years. In addition, it provides information on the overall cancer mortality as well as potential harms as a consequence of overdiagnosis and overtreatment due to false positives. Such fact boxes enable health-care consumers to make more informed decisions for or against treatments based on the currently available scientific evidence.

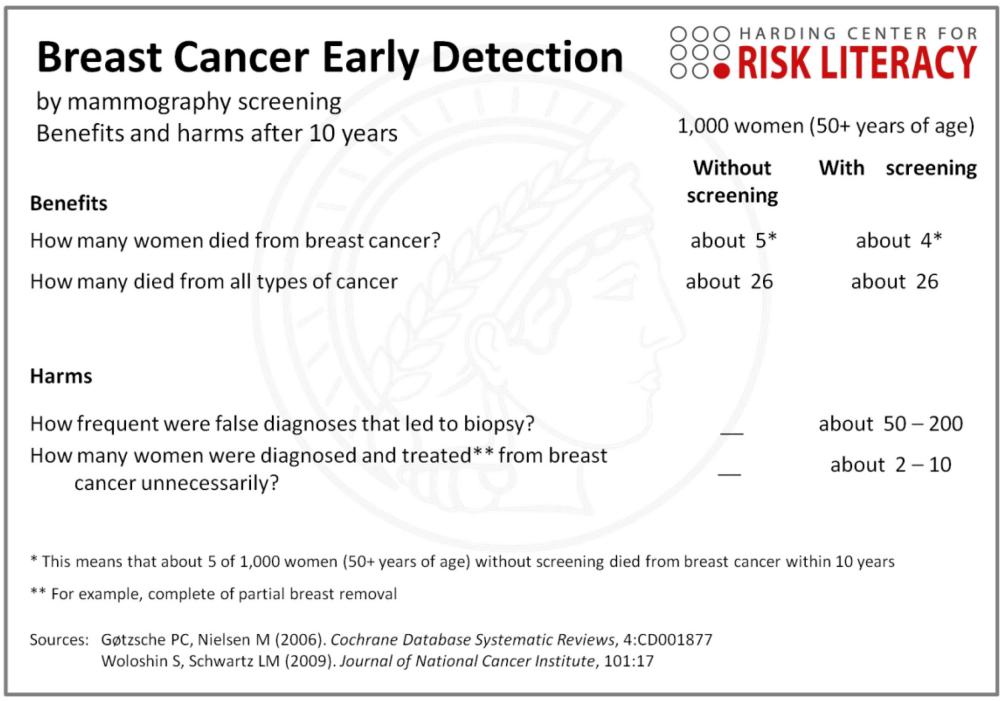


Figure 8. Fact box summarizing medical evidence of the effects of mammography on breast cancer mortality and cancer mortality. In addition, information about potential harms (e.g., overdiagnosis and overtreatment) is communicated (Source: www.harding-center.com/fact-boxes/mammography, retrieved on 15 March 2012).

Author Note

This research was supported by the joint program “Acciones Integradas Hispano-Alemanas” from the DAAD and the Ministerio de Ciencia y Tecnología. BM was supported by Grant ME 3717/2–1 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program “New Frameworks of Rationality" (SPP 1516). Portions of this research were presented at TeaP 2012 in Mannheim, Germany. We thank Rona Unrau for editing the paper.

References

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., Costiniuk, C., et al. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews (Online)*, (3), CD006776.

Arkes, H. J., & Gaissmaier, W. (in press). Psychological research and the Prostate-Cancer Screening controversy. *Psychological Science*.

Barratt, A. (2004). Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *Canadian Medical Association Journal*, *171*, 353–358.

Bodemer, N., & Gaissmaier, W. (2012). Risk communication in health. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson (Eds.), *Handbook of Risk Theory* (pp. 621–660). Dordrecht: Springer Netherlands.

Bodemer, N., Müller, S.M., Okan, Y., Garcia-Retamero, R., & Neumeyer-Gromen, A. (2012) Do the media provide transparent health information? A cross-cultural comparison of publicinformation about the HPV vaccine. *Vaccine*, *30*, 3747–3756.

Brase, G. L. (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review*, *15*, 284–289.

CDC's Diabetes Program - Data & Trends - *Prevalence of Cardiovascular Disease - Age-Adjusted Percentage of Persons with Diabetes Aged 35 Years and Older Reporting Any Cardiovascular Disease Condition, by Sex*, *United States, 1997–2009*. Retrieved from, http://www.cdc.gov/diabetes/statistics/cvd/fig5.htm

Collins, R., Armitage, J., Parish, S., Sleight, P., & Peto, R. (2004). Effects of cholesterol-lowering with simvastatin on stroke and other major vascular events in 20536 people with cerebrovascular disease or other high-risk conditions. *Lancet*, *363*, 757–767.

Cosmides, L., & Tooby, J. (1995). Are humans good intuitive statisticians after all?: Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Covey, J. (2007). A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making*, *27*, 638–654.

Cranney, M., & Walley, T. (1996). Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *The British Journal of General Practice*, *46*, 661–663.

Edwards, A., Elwyn, G., Covey, J., Matthews, E., & Pill, R. (2001). Presenting risk information–a review of the effects of „framing“ and other manipulations on patient outcomes. *Journal of Health Communication*, *6*, 61–82.

Forrow, L., Taylor, W. C., & Arnold, R. M. (1992). Absolutely relative: How research results are summarized can affect treatment decisions. *The American Journal of Medicine*, *92*, 121–124.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*, 210–216.

Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, *170*, 462–468. doi:10.1001/archinternmed.2009.481

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American Journal of Public Health*, *99*, 2196–2202. doi:10.2105/AJPH.2009.160234

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: overcoming challenges in people's understanding of risks. *Social Science & Medicine*, *70*, 1019–1025.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*, 53–96.

Gigerenzer, G., & Gray, M. (2011) Launching the century of the patient. In G. Gigerenzer & M. Gray (Eds.) *Better doctors, better patients, better decisions: Envisioning health care 2020* (pp. 3–28). Cambridge, MA: MIT Press.

Gigerenzer, G., Wegwarth, O., & Feufel, M. (2010). Misleading communication of risk. *British Medical Journal*, *341*, c4830.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Gøtzsche, P. C., & Nielsen, M. (2006). Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews (Online)*, (4), CD001877.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Higgins, J.P.T., Green, S. (2009). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Retrieved from www.cochrane-handbook.org.

Jorgensen, K. J. (2004). Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study. *British Medical Journal*, *328*, 148–150.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44.

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, *8*, 543–548.

Moser, C. (2002). *Improving Numeracy and Literacy: A Fresh Start*. Nottingham: DFES Publications.

Moynihan, R., Bero, L., Ross-Degnan, D., Henry, D., Lee, K., Watkins, J., Mah, C., & Soumerai, S. (2000). Coverage by the news media of the benefits and risks of medications. *The New England Journal of Medicine*, *22*, 1645–1650.

Mühlhauser, I., Kasper, J., Meyer, G., & Federation of European Nurses in Diabetes. (2006). Understanding of diabetes prevention studies: questionnaire survey of professionals in diabetes care. *Diabetologia*, *49*, 1742–1746.

Peters, E. (2008). Numeracy and the perception and communication of risk. *Annals of the New York Academy of Science*, *1128*, 234–267.

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

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

Sarfati, D., Howden-Chapman, P., Woodward, A., & Salmond, C. (1998). Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. *Journal of Medical Screening*, *5*, 137–140.

Schwartz, L.M., Woloshin, S., Dvorin, E.L., & Welch, H.G. (2006). Ratio measures in leading medical journals: Structured review of accessibility of underlying absolute risks. *British Medical Journal*, *333*, 1248–1252.

Schwartz, L. M., Woloshin, S., & Welch, H. G. (2009). Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Annals of Internal Medicine*, *150*, 516–527.

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

Sedrakyan, A., & Shih, C. (2007). Improving depiction of benefits and harms: Analyses of studies of well-known therapeutics and review of high-impact medical journals. *Medical Care*, *45*, 23–28.

Slater, M. D., Long, M., Bettinghaus, E. P., & Reineke, J. B. (2008). News coverage of cancer in the United States: a national sample of newspapers, television, and magazines. *Journal of Health Communication*, *13*, 523–537.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315. doi:10.1037/0033-295X.90.4.293

Williams, D., Kelly, A., Carvalho, M., & Feely, J. (1998). Effect of the British warning on contraceptive use in the General Medical Service in Ireland. *Irish Medical Journal*, *91*, 202–203.

Appendix

Scenario used in Experiment 2a (Relative risk reduction)

A study tested a new drug for diabetes. The study's aim was to find out whether and to what extent the new drug reduced the risk of heart disease. To evaluate the new drug, 2,000 patients with diabetes were tested. The patients were randomly assigned to two groups: 1,000 patients received the new drug and 1,000 patients received no drug. The patients receiving no drug served as a control group. After five years, the number of patients with heart diseases in each of the two groups was compared.

Subjects in the *baseline percentage condition* received then the following information:

In the control group without the drug 30% of the patients had heart disease.
The study showed that the new drug reduced the risk of heart disease by 20%.
Please estimate how many patients in the group with the drug suffered from heart disease:

O 24% out of 1,000 patients who received the drug had heart disease
O 10% out of 1,000 patients who received the drug had heart disease
O 20% out of 1,000 patients who received the drug had heart disease
O 28% out of 1,000 patients who received the drug had heart disease
O 2% out of 1,000 patients who received the drug had heart disease

Subjects in the *baseline frequency condition* received the same information, but here the baseline risk was presented in frequency format:

In the control group without the drug 300 out of 1,000 of the patients had heart disease.
The study showed that the new drug reduced the risk of heart disease by 20%.
Please estimate how many patients in the group with the drug suffered from heart disease:

O 240 out of 1,000 patients who received the drug had heart disease
O 100 out of 1,000 patients who received the drug had heart disease
O 200 out of 1,000 patients who received the drug had heart disease
O 280 out of 1,000 patients who received the drug had heart disease
O 20 out of 1,000 patients who received the drug had heart disease

Scenario used in Experiment 2b (Relative risk increase)

A study tested a new drug for patients with diabetes. The study's aim was to find out whether and to what extent the new drug increased the risk of heart attack (a side-effect of the drug) in patients with diabetes. To evaluate the new drug's side-effects, 2,000 patients with diabetes were tested. The patients were randomly assigned to two groups: 1,000 patients received the new drug and 1,000 patients received no drug. The patients receiving no drug served as a control group. After five years, the number of patients with heart attack in each of the two groups was compared.

Subjects in the *baseline percentage condition* received then the following information:

In the control group without the drug 30% of the patients had heart disease.
The study showed that the new drug increased the risk of heart disease by 20%.
Please estimate how many patients in the group with the drug suffered from heart disease:

O 36% out of 1,000 patients who received the drug had heart disease
O 50% out of 1,000 patients who received the drug had heart disease
O 20% out of 1,000 patients who received the drug had heart disease
O 32% out of 1,000 patients who received the drug had heart disease
O 2% out of 1,000 patients who received the drug had heart disease

Subjects in the *baseline frequency condition* received the same information, but here the baseline risk was presented in frequency format:

In the control group without the drug 300 out of 1,000 of the patients had heart disease.
The study showed that the new drug reduced the risk of heart disease by 20%.
Please estimate how many patients in the group with the drug suffered from heart disease:

O 360 out of 1,000 patients who received the drug had heart disease
O 500 out of 1,000 patients who received the drug had heart disease
O 200 out of 1,000 patients who received the drug had heart disease
O 320 out of 1,000 patients who received the drug had heart disease
O 20 out of 1,000 patients who received the drug had heart disease

Chapter 5

Providing ambiguous health information: Effects on treatment choice and choice strategies

Nicolai Bodemer[1, 2], Stephanie M. Müller[3], & Markus A., Feufel[1, 2]

[1]Harding Center for Risk Literacy, Berlin, Germany
[2]Max Planck Institute for Human Development, Berlin, Germany

[3]Center for Empirical Research in Economics and Behavioral Science (CEREB), University of Erfurt, Erfurt, Germany

[Word count: 7866]

Bodemer, N., Müller, S.M., & Feufel, M.A. (2012). *Providing ambiguous health information: Effects on treatment choice and choice strategies*. Manuscript in preparation.

Abstract

Ambiguity is inherent in information about benefits and harms of medical treatments, but it is usually not included in health risk communication. By ambiguity we mean imprecise estimates about treatment effects such as ranges, confidence intervals, or standard errors. Research on monetary gambles found that people are ambiguity averse—they prefer certain over ambiguous options—in the domain of gains, but ambiguity seeking—they prefer ambiguous over certain options—in the domain of losses. We aimed at transferring the concepts of ambiguity to medical treatment choice and how people choose between to treatments that differ in the degree of ambiguity and their average rate of benefits or harms. In Study 1, we presented participants with two treatments that offered either certain or ambiguous information. Treatment options were either presented as benefits or harms. We found that (1) participants were able to identify superior options, even when ambiguity was present and (2) did not show ambiguity aversion in the benefits condition or ambiguity seeking in the harms condition. Experiment 2 extended these findings by mapping participants' choice strategies. Results suggest high heterogeneity in peoples' choice strategies. Neither ambiguity aversion nor ambiguity seeking can sufficiently explain choice strategies. Large proportions of participants focused primarily on either the lower or upper bounds of the treatment options, or compared treatments average rates by focusing on the range's midpoint. We suggest that including ambiguity in health risk communication allows people to make individual and informed decisions based on evidence and its limitations.

Introduction

Uncertainty plays an important role in medicine (Politi, Col, & Han, 2007). First, scientific evidence is limited. Even randomized controlled trials – often regarded as the gold standard in medical research – have limitations due to design principles, sample size, and lack of validity and reliability of measures. Second, risk estimates are based on population data and therefore cannot be applied one-to-one to individuals. Third, risk estimates are based on past events. Their application to the present and future rests on the assumption that the environment and underlying forces do not change. In this paper, we refer to a particular type of uncertainty – ambiguity. Ambiguity describes the imprecision of an estimate due to limited reliability, credibility, or adequacy of information (Camerer & Weber, 1992) or—in other words— "uncertainty about probability, created by missing information" (Frisch & Baron 1986). Expressions of ambiguity are probabilistic parameters such as standard deviations, expert confidence ratings, or ranges (Han, Klein, & Arora, 2011; Politi et al., 2007). For instance, a patient's lifetime risk for colorectal cancer can be expressed as a precise point estimate such as 9%, or as an ambiguous estimate of between 5%-13% (Han, Klein, Lehman, Killam, Massett, & Freedman, 2010). Moreover, ambiguity here refers to scientific-centered (data-centered) ambiguity; other sources comprise system- and patient-centered ambiguity which are not part of this study (Han et al., 2011).

The inclusion of ambiguity in medical risk communication conforms to the ideal of shared decision making—to present transparent and complete information as basis to include patients in their decisions about health (Edwards & Elywn, 2001; Gigerenzer & Gray, 2011). Medical associations such as the Cochrane Collaboration recommend to include uncertainty information in form of confidence intervals (with exact p-values) to evaluate the quality of the available evidence and "assess the clinical usefulness of the intervention" (Higgins & Green, 2009, p.12.10). However indicators of uncertainty are rarely, if ever, included in patient

information (e.g., Bodemer, Müller, Okan, Garcia-Retamero, & Neumeyer-Gromen, 2012; Politi et al., 2007) or in physician-patient dialogues (Braddock, Edwards, Hasenberg, Laidley, & Levinson, 1999).

Whereas some argue for disclosing ambiguity or uncertainty to provide a more complete picture of the quality of the available evidence (Ancker, Senathirajah, Kukafka, & Starren, 2006; Feufel, Antes, & Gigerenzer, 2010; Han et al., 2009), others voice major objections. Some experts believe that patients are incapable of understanding and reasoning about uncertainty. For instance, when food experts were asked about the disclosure of uncertainty about food risks to the public, most experts believed that the public cannot conceptualize uncertainty and would react with panic and confusion (Frewer, Hunt, Brennan, Kuznesof, Ness, & Ritson, 2003). This view is supported by research showing that people have difficulty in processing ambiguous information (e.g., Epstein, 1999; Viscusi, Wesley & Huber, 1991) and tend to maintain an illusion of certainty (Gigerenzer, 2003; Gigerenzer, Gaissmaier, Kurz-Mielcke, Schwartz, & Woloshin, 2007).

Communication and effects of ambiguity

Due to unfamiliarity with the concepts of ambiguity (Ibrekk & Morgan, 1988; Johnson & Slovic, 1995), people's behavior is often described as ambiguity averse (Ellsberg, 1961). Ambiguity aversion[2] describes the phenomenon that people tend to prefer known risks (i.e., certain options) over unknown risks (i.e., ambiguous options) in monetary gambles (Curley & Yates, 1985; Einhorn & Hogarth, 1986; Epstein, 1999; Keren & Gerritsen, 1999; Camerer & Weber, 1992). For instance, participants are willing to pay a higher price in a gamble with a precise probability compared with an ambiguous probability (Fox & Tversky, 1995). Even when the odds favor the ambiguous treatment, some researchers still found that people are

[2] We will use the term ambiguity aversion here (see also Han et al., 2011; Politi et al., 2007), but it should be noted that it is closely related to the concept of risk aversion which describes that people prefer certain payoffs over uncertain payoffs (see also Epstein, 1999).

ambiguity averse and prefer the certain but less favorable option (Keren & Gerritsen, 1999). Whether a decision maker is ambiguity averse depends on whether gains or losses are at stake. Ambiguity aversion has been primarily observed in the domain of gains, while changing the perspective to losses shifts people's tendency of ambiguity aversion to ambiguity seeking (Laughhunn, Payne, & Crum, 1980). What underlies people's tendency of ambiguity aversion? Heath and Tversky (1991) argued that ambiguity generally makes people feel uncomfortable as it signals a lack of knowledge. Other research argues that presenting a range of possible outcomes may shift attention to the worst outcome which people then try to avoid (Einhorn & Hogarth, 1985, Visuci et al., 1991). As we will show later this does not necessarily imply to avoid ambiguous options.

To date, only little is known about the influence of ambiguity in medical risk information on risk perception and medical choice behavior. In medicine, presenting health risks in terms of a range of possible outcomes (e.g., 5% - 13% lifetime risk of colon cancer) instead of as a point estimate (e.g., 9% lifetime risk of colon cancer) increases perceived risk, worry and distress among patients (Einhorn & Hogarth, 1985; Han et al., 2009, 2010). Moreover, some researchers argue that patients are similarly ambiguity averse as in monetary gambles and point to results that show people's decreased engagement in particular health behaviors such as willingness to participate in PSA screening for early detection of prostate cancer when presented with ambiguous information (Briss, 2004; Frosch, Kaplan, & Felitti, 2001; Raffle, 2001). Yet, other researchers did not find a direct effect of ambiguity on health behaviors such as, for instance, uptake rates for early detection of cancer (Farrell, Murphy, & Schneider, 2002; Taplin, Urban, Taylor, & Savarino, 1997). Like for monetary gambles, the influence of ambiguity might change depending on whether ambiguity relates to medical benefits or harms. The benefits of a treatment resemble a gain situation and may lead to ambiguity aversion. In contrast, the harms of a treatment resemble a loss situation and people

may react with ambiguity seeking. To the best of our knowledge, research has not yet addressed this question in medical treatment choice.

In this research, we aimed at investigating the role of ambiguity, and in particular whether ambiguity aversion and seeking found in monetary gambles exist also in medical choices. Thereby, we focused on answering the following questions. First: Do people react averse to ambiguity when choosing medical treatments when benefits are presented? Do people seek ambiguity when choosing medical treatments when harms are presented? Second: Do people identify a superior[3] treatment despite the presence of ambiguity? How do people choose when such a superior treatment is not present? Third: Which choice strategies underlie treatment selection?

The influence of numeracy on decision making under ambiguity

A growing body of literature highlights the importance of individual differences in decision making tasks (Appelt, Milch, Handgraaf, & Weber, 2011). When it comes to people's understanding of health risks, a cognitive ability that has been found to predict choices is numeracy (Lipkus, Samsa, & Rimer, 2001; Reyna, Nelson, Han, & Dieckmann, 2009). High numerates have been found to be more precise in their numerical interpretation and less prone to framing effects (Garcia-Retamero & Galesic, 2010; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006; Peters & Levin, 2008). Moreover, low numerates have difficulties in choosing normatively better options and choose based on different strategies compared with high numerates (Pachur & Galesic, 2012). In a similar vein, research on the perceived risk of breast cancer documents that more educated women appreciate ambiguity information in form of ranges whereas less educated women interpret a range as vague and

[3] Here, the superior option always refers to a higher rate of benefits or lower rate of harms on any of the values independent of whether one compares the average rates of the treatments or the lower and upper bound of the ranges.

confusing (Schapira, Nattinger, & McHorney, 2001). We investigate whether numeracy moderates effects of ambiguity on people's treatment choice.

Hypotheses

In our experiments, participants had to choose between two treatment options that differed in their degree of ambiguity. One option's effectiveness was certain (presented as point estimate of rate of benefits/ harms), and one option's effectiveness was ambiguous (range of benefits/ harms). We used a range to communicate ambiguity in line with previous studies on ambiguity in risk communication (e.g., Han et al., 2009; Han et al., 2010; Schapira et al., 2001). Besides differences in the degree of ambiguity, treatments had either equivalent average rates of benefits or harms, or differed in the average rates with one having a higher average rate. The average rate of the certain option corresponded to the point estimate; the average rate of the ambiguity option corresponded to the midpoint of the range (Han et al., 2009; Schapira et al., 2001). Finally, either benefits or harms were presented. We hypothesized the following in the different conditions:

1. *When treatments differ in their degree of ambiguity, but not in their average rates*

   H1a: The majority of participants shows ambiguity aversion and avoids the ambiguous option when benefits are presented.

   H1b: The majority of participants seeks ambiguity and chooses the ambiguous option when harms are presented.

2. *When treatments differ in their degree of ambiguity and average rates*

   H2a: The majority of participants indentifies and chooses the superior option in the benefits and harms condition, independently of ambiguity.

   H2b: The majority of participants choose the option with higher average rate when no clear superior option is presented in the benefits and harms condition.

H3: Low numerates have more difficulty in finding a superior option, so we expect to observe higher variance in choices in low numerates compared with high numerates.

Study 1

Experiment 1a

Method

Design and Procedure. To avoid any influence of prior knowledge, participants received the following hypothetical medical scenario.

In this study, we ask you to imagine that you have been feeling sick for two days with fever, headaches, and fatigue. Your doctor diagnoses that you have a bacterial infection called TIRA. If left untreated, the symptoms will persist for at least 4 weeks and the risk of relapses will be elevated.

We will present you with four different scenarios that provide you with information about the treatment for TIRA. The **benefits** of the treatments **are similar for all scenarios**, whereas **the harms differ**. Evidence of the potential **harms is displayed in a red box.**

Please read the information carefully and answer each question.

Note: We are interested in your personal opinion, there is no right or wrong answer!

Subsequently, participants received six scenarios in random order. Between subjects, we presented participants either with treatment benefits (number of people who are symptom-free after 3 days) or treatment harms (number of people who experience stomach bleeding as side effect); everything else was held constant. This manipulation aimed to elicit a gain or loss situation. In each scenario, participants had to choose between two medical treatments presented as a certain option (e.g., 20 out of 100 patients taking this treatment are symptom free after 3 days) and an ambiguous option (e.g., between 10 – 30 out of 100 patients taking this treatment are symptom free after 3 days). The six scenarios can be grouped into three sets of two scenarios each (see Table 1).

Set I: Scenarios A and B each showed two treatments with the same average rate of benefits or harms, but one option being certain and one ambiguous. Ambiguity was high

(range = 20) in scenario A and low (range = 4) in scenario B. The objective was to demonstrate whether participants prefer the certain over the ambiguous option when their average rates of benefits and harms are identical.

Set II: In scenarios C and D, treatment 1 had higher rate of benefits or harms than treatment 2 (20 out of 100 vs. 17 out of 100). In scenario C, the option with the higher average rate (treatment 1) was presented as certain and the one with the lower average rate (treatment 2) with low ambiguity (range = 4). In scenario D, treatment 1 had low ambiguity, whereas treatment 2 was certain. In both scenarios, treatment 1 was clearly superior to treatment 2 in the benefits condition, and treatment 2 superior to treatment 1 in the harms condition. This set aimed at testing whether participants were able to identify the superior treatment option when presented with ambiguous information.

Set III: In scenarios E and F, treatment 1 again had a higher average rate than treatment 2 (20 out of 100 vs. 17 out of 100). In scenario E, treatment 1 was certain and treatment 2 highly ambiguous (range = 20); this manipulation was identical but reversed in scenario F. This set aimed at investigating participants' choice behavior when ranges and point estimate overlap and an overall superior option was absent.

The order of the scenarios and the two treatments was randomized within the scenarios. After all six scenarios, we assessed participants' numeracy skills using the Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012).

Participants. A total of 106 participants of which 54% were female took part in this experiment (n=54 in the benefits and n=52 in the harms condition). The average age was 33 years (SD=12.1). Participants were recruited via Amazon MTurk and randomly assigned to one of two conditions {harms, benefits}. Remuneration was $0.75.

Table 1

The Three Sets and Six Scenarios of Experiment 1a. Random Halves of Participants Were Told That the Treatment Information Is Either About Benefits or Harms, Respectively. In The Benefits Condition, a Treatment Is Better the Higher the Rates Are, Whereas in the Harms Condition Lower Values Signal a Better Treatment.

Set	Ambiguity	Scenario	Treatment 1	Treatment 2
I (same average rates, high or low ambiguity)	certain vs. high	A	20	10 – 30
	certain vs. low	B	20	18 – 22
II (different average rates, low ambiguity)	certain vs. low	C	20	15 – 19
	low vs. certain	D	18 - 22	17
III (different average rates, high ambiguity)	certain vs. high	E	20	7 – 27
	high vs. certain	F	10 – 30	17

Results

Treatment Choice. We first present the results for the benefit condition (Figure 1). In set I, when treatments differed on their degree of ambiguity (certain vs. ambiguous), more people chose the certain option when ambiguity was high (scenario A: 65% vs. 35%, two-tail

binomial test, *p=.04*). However, when ambiguity was low, only about half of the participants chose the certain option (scenario B: 56% vs. 44%, two-tail binomial, *p =.49*). Hence, on an aggregated level, participants did not show ambiguity aversion. In set II—different average rates, low ambiguity (range = 4)—the majority of participants chose the treatment with the higher average rate (91% and 94%), that is, the option that was superior (two-tail binomial, *p=.001*). In other words, participants had no difficulty in identifying the superior option despite ambiguity. In set III— different average rates, high ambiguity (range = 20)—we found that the majority chose the option with the higher average rate (91%, *p=.001*) when it was presented as certain (scenario E). However, only 72% (*p=.001*) chose this option when presented with high ambiguity (scenario F); that implies that 28% chose the certain estimate that was on average inferior. Hence, one fourth reacted in line with the concept of ambiguity aversion.

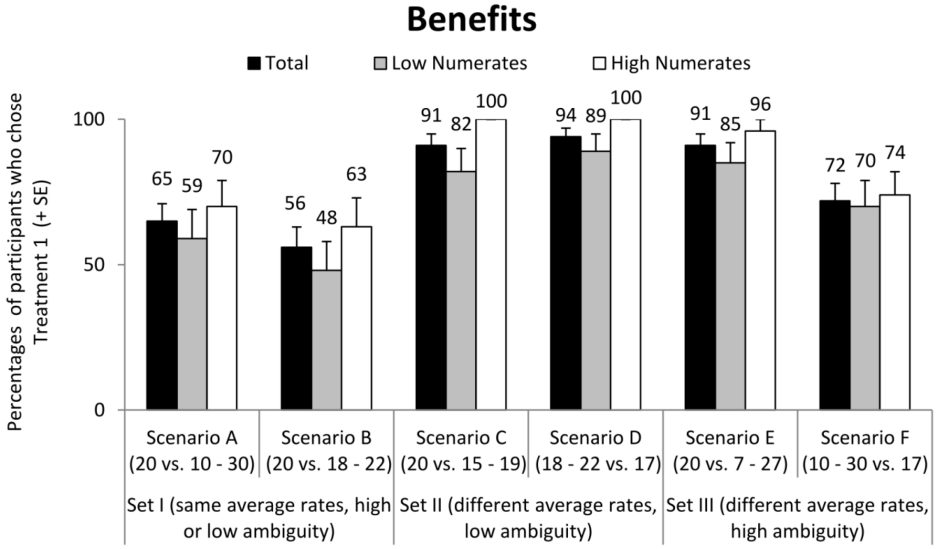


Figure 1. Percentages of participants who chose treatment 1 in Experiment 1a for the benefits condition. Total refers to all participants (*n*=54); low numerates to participants with a numeracy score <2 (*n*=27); high numerates to participants with a score ≥2 (*n*=27).

When the outcomes were presented as harms (Figure 2), lower values (lower average rates) signaled a better treatment (i.e., the fewer out of 100 patients experience stomach bleeding). In set I, 56% in scenario A chose the treatment with high ambiguity (treatment 2) and 61% in scenario B chose the treatment with low ambiguity (two-tailed binomial, *p=.488*; *p=.126*). Hence, half of the participants made choices consistent with ambiguity seeking. In set II, all participants chose treatment 2—the one with the lower average rate of harms—over treatment 1 (in scenario C and D), independently of whether it was certain or had low ambiguity. More precisely, participants identified the superior option. In set III, when one option was certain and one had high ambiguity, 85% of participants chose treatment 2 (the one with a lower average rate of harms) when it was (highly) ambiguous (two-tailed binomial, *p=.001)*, while 75% did so when it was certain. In other words, one fourth chose consistently with the prediction of ambiguity seeking in scenario F (two-tailed binomial, *p=.002)*.

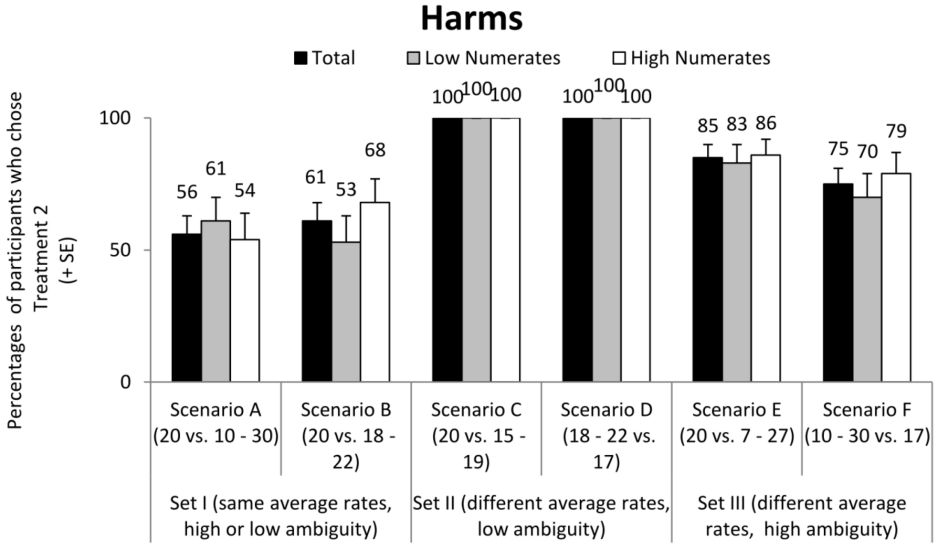


Figure 2. Percentages of participants who chose treatment 2 in Experiment 1a for the harms condition. Total refers to all participants (*n*=52); low numerates to participants with a numeracy score <2 (n=23); high numerates to participants with a score ≥2 (*n*=28).

Numeracy. We conducted a median split to group participants into low (numeracy score <2, 47% of participants) and high numerates (numeracy score ≥2, low: 47%, high: 52%, missing: 1%). In the benefits condition, in set I, 59% of the low numerates in scenario A and 56% in scenario B chose the certain option compared with 70% and 63% of the high numerates. In set II, all high numerates chose the superior option in both scenarios. Among low numerates, 82% and 89% chose this option (for scenarios C and D, respectively). In set III, 85% of low numerates and 96% of high numerates chose the on average better option when it was certain (scenario E). In scenario F—where the on average better option had high ambiguity—70% of the low numerates and 74% of the high numerates chose the on average better and high ambiguity option.

In the harms condition, in set I, 61% of low numerates compared with 54% of high numerates chose the high ambiguity option (scenario A). In scenario B, 53% of the low numerates and only 68% of high numerates chose the low ambiguity option. In set II, independent of numeracy, all participants chose treatment 2. Finally, in set III, scenario E showed an equal percentage of low and high numerates choosing treatment 2 (83% and 86%); in scenario F, 70% of low compared with 79% of high numerates chose treatment 2.

Summary of Experiment 1a

In summary, when two treatments had the same average rate, but differed in their degree of ambiguity, we found a slight tendency for ambiguity aversion in the benefits condition when ambiguity was high. When treatments differed also in the average rate, the vast majority chose the superior option when ambiguity was low. When ambiguity was high, choices were influenced by the degree of ambiguity. In the benefits condition, when the on average better option had high ambiguity, about one fourth of participants chose the certain but on average inferior option. In the harms condition, when the on average inferior option had high ambiguity, about one fourth of participants chose this option over the on average

better and certain option. Across all scenarios, choices of low numerates showed higher variance.

Experiment 1b

In Experiment 1a one treatment option was presented as certain and the other one as ambiguous. Do participants show a similar choice behavior, when both treatments are ambiguous? Participants who were presented with one certain and one ambiguous treatment option may have doubted the precision of the certain option. Consequently, they may have formed some "subjective" ambiguity around the precise estimate and also interpreted it as the midpoint of an unspecified range. Moreover, as pointed out in the introduction, a precise point estimate can rarely be justified in evidence based medicine (Politi et al., 2007). Hence, Experiment 1b aimed at studying participants' choice behavior when both treatment options include ambiguity information.

We expected similar results as in Experiment 1a—the substitution of the certain option with a low ambiguous option should not change people's choice pattern. More precisely, we expected participants to be indifferent when only the degree of ambiguity between options differed. When options differ in their average outcomes, we expect that participants can identify the overall superior option. When there is no superior option—e.g., due to different ranges—we still expect that the majority of participants choose the option with the higher average rate in the benefits condition, and the lower average rate in the harms condition.

Method

Design and Procedure. We used the same design as in Experiment 1a with one modification. Instead of presenting one treatment as certain and the other as ambiguous, both treatments were ambiguous but differed in their degree of ambiguity (low vs. high). The number of scenarios was thereby reduced to four. Table 2 gives an overview of the four scenarios and the three sets. Again, set I represented a scenario in which treatments have the

same average rate but one option had low ambiguity (range = 4) and the other one high ambiguity (range = 20). In set II, both treatments were high in ambiguity (range = 20) and differed in their average rate, which was higher in treatment 1. Finally, in set III, treatments differed in both the average rate and the degree of ambiguity. For instance, treatment 1 in scenario C had a higher average rate than treatment 2 and low ambiguity, whereas treatment 2 had a lower average rate and high ambiguity. In scenario D, treatment 1 had a higher average rate than treatment 2 and high ambiguity, whereas treatment 2 had a lower average rate and low ambiguity.

Table 2.

The Four Scenarios of Experiment 1b. Random Halves of Participants Were Told that the Treatment Information Is Either About Benefits or Harms, Respectively.

Set	Ambiguity	Scenario	Treatment 1	Treatment 2
I (same average rates, low vs. high ambiguity)	low vs. high	A	18 - 22	10 – 30
II (different average rates, high ambiguity)	high vs. high	B	10 - 30	7 - 27
II (different average rates, low vs. high ambiguity)	low vs. high	C	18 - 22	7 – 27
	high vs. low	D	10 – 30	15 – 19

Again, half of participants received scenarios with the information about treatment benefits; the others received scenarios with the information about treatment harms. Numeracy was assessed after participants had made their choices in all four scenarios.

Participants. In total, 104 participants took part in this Experiment (*n*=53 in the benefits and *n*=51 in the harms condition). The average age was 36 years (*SD*=12.6), and 55% were female. Participants were recruited via Amazon MTurk and randomly assigned to one of the two conditions {harms, benefits}. Remuneration was $0.75.

Results

Treatment Choice. First, we present results for the benefits condition (Figure 3). In set I (same average rates, low vs. high ambiguity), 57% chose the treatment with low ambiguity (two-tail binomial test, *p=.41*). Similarly to Experiment 1a, we did not find an effect of ambiguity aversion. In set II (different average rates, high ambiguity), the vast majority (91%) chose the superior option (treatment 1, two-tail binomial, *p=.001*). In set III (different average rates, low vs. high ambiguity), similar to Experiment 1a, the majority in scenario C chose treatment 1 over treatment 2 (93%; two-tail binomial, *p=.001*); in scenario D 72% chose treatment 1—consequently, one fourth chose the on average inferior, but less ambiguous treatment (two-tail binomial, *p=.002*).

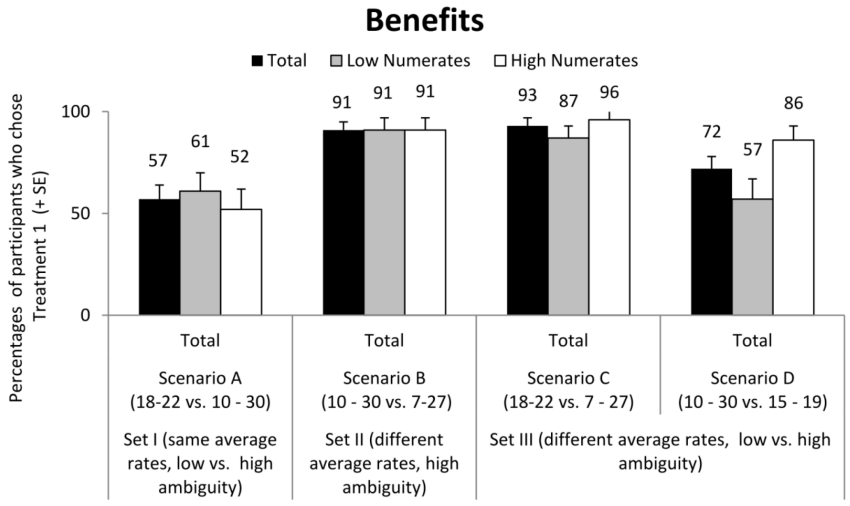


Figure 3. Percentages of participants who chose treatment 1 in Experiment 1b for the benefits condition. Total refers to all participants (*n*=53); low numerates to participants with a numeracy score <2 (*n*=23); high numerates to participants with a score ≥2 (*n*=29).

When it comes to harms (Figure 4), about half of the participants chose the less ambiguous option (47% vs. 53% choosing the high ambiguous option; two-tail binomial, *p=.78*). In set II, the vast majority (94%, two-tailed binomial, *p=.001*) chose treatment 2 that was superior (i.e., it showed a lower rate of harms). In set III, again similar to Experiment 1a, choices were influenced by the degree of ambiguity. In scenario C, most participants chose treatment 2 (91%, two-tailed binomial *p=.001*), but in Scenario D only 81% (two-tailed binomial, *p=.001*) chose treatment 2 that had a lower average rate and lower ambiguity.

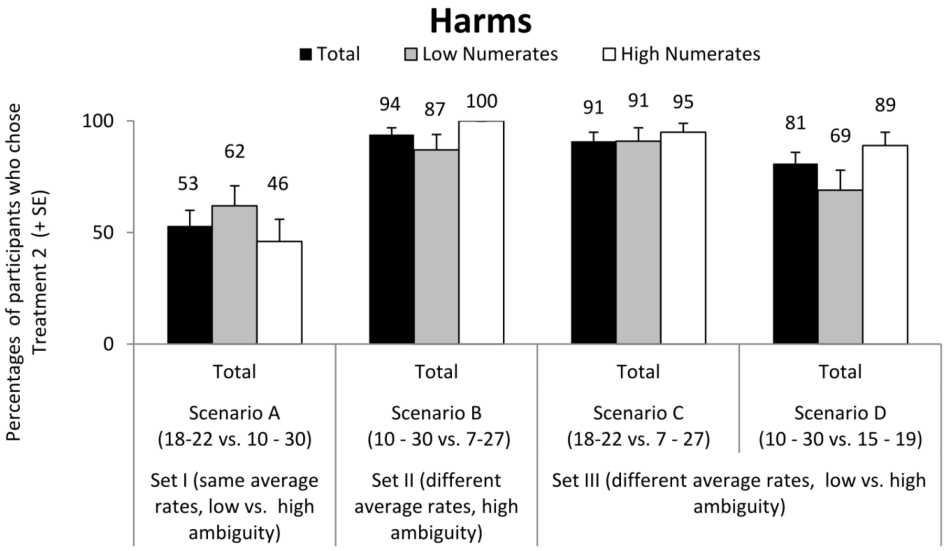


Figure 4. Percentages of participants who chose treatment 2 in Experiment 1b for the harms condition. Total refers to all participants (n=51); low numerates to participants with a numeracy score <2 (n=16); high numerates to participants with a score ≥2 (n=35).

Numeracy. Again, we categorized participants as low numerates and high numerates (low 39%, high: 58%, missing: 3%). In the benefits and harms condition, across all scenarios, low numerates showed higher variance in choices than high numerates (see Figure 3, 4). Particularly in scenario D in the benefits condition, 57% of low numerates versus 86% of high

numerates chose treatment 1. In the harms condition of scenario D, 69% of low and 89% of high numerates chose treatment 2.

Summary of Experiment 1b

In summary, results of Experiment 1b extended and replicated findings of Experiment 1a. When both treatments had the same average rates and only differed in the degree of ambiguity, we did not find a majority of participants being ambiguity averse in the benefits condition or ambiguity seeking in the harms condition. Moreover, when both treatments only differed in the average rate and had equal ambiguity, participants chose the superior option. Again, there was a difference between the harms and benefits condition. In the benefits condition, variance in choice was higher when the on average better option had high ambiguity. Hence, more participants preferred the certain option, although it was on average inferior. In the harms condition, the pattern was the opposite: when the on average inferior option had high ambiguity, more people choose this option instead of the on average superior and less ambiguous option.

Discussion Study 1

We did not find strong evidence supporting that people are ambiguity averse when presented with treatment benefits (H1a), or ambiguity seeking when presented with treatment harms (H1b). Only in the benefits condition when one option was certain and one highly ambiguous, a small majority chose the certain option. Yet, when two treatments differed only in their degree of ambiguity, we did not observe a majority opting for the certain option in the benefits condition, or the ambiguous option in the harms condition. Moreover, even when presented with ambiguity, participants were able to find a superior option, independent of whether benefits or harms were presented (H2a). When there was no superior option, we still found that the majority chose the treatment that had the higher average rate of benefits or the lower rate of harms (H2b). However, up to one third of participants in the benefits condition

chose the option with the lower average when this was certain. This indicates that at least a minority acted in line with ambiguity aversion. In the harms condition, about one fourth of participants chose the option with the higher average rate of harms when this had high ambiguity. This indicates that at least a minority sought for ambiguity in the harms condition, most likely because the option with high ambiguity had a smaller lower bound. Hence, despite of the worse average rate and greater upper bound, some participants were rather optimistic by focusing on the lowest possible outcome.

Across scenarios, we observed a similar pattern for low and high numerates (H3). Hence, low and high numerates did neither differ with respect to ambiguity aversion in the benefits scenarios nor ambiguity seeking in the harms scenarios. Moreover, both groups were able to identify a superior option if present. Yet, low numerates showed slightly higher variance in choices, so primarily low numerates accounted for differences in treatment choice.

Ambiguity aversion and ambiguity seeking are only one potentially explanation for people's choices under ambiguity. Heath and Tversky (1991) propose that people react averse towards ambiguity because they feel uncomfortable with the lack of knowledge. Viscusi and colleagues (1991) propose that ambiguity shifts people's attention to the worst possible outcome. In our study, the question why some participants selected the certain option over the ambiguous option in the benefits condition could also be a consequence of the lower bound of the range which was smaller when ambiguity was high. In other words, these participants could have also focused on the worst possible outcome (the lowest possible rate of benefits) of each option. Likewise, in the harms condition, those opting for the treatment with high ambiguity must have focused on the lower bound—the best possible outcome (the lowest possible rate of harms) which coincided with the high ambiguity option. Experiment 1a/b did not allow us to disentangle the actual reason why a participant chose a particular treatment—it

could have been due to ambiguity aversion, the lower bound, the upper bound, or the midpoint of the range of possible benefits of harms.

Study 2

The main objective of Study 2 was to map peoples' underlying choice strategies. As previously pointed out, ambiguity aversion is only one potential explanation (Heath & Tversky, 1991). Instead of focusing on the width of the range that signals the degree of ambiguity, one might shift attention to one of the bounds—either to avoid the worst possible outcome (Einhorn & Hogarth, 1985, Visuci et al., 1991), or to seek the best possible outcome. For example, when comparing two treatments' benefits, a patient might be primarily interested in avoiding the option that has a lower possible rate of benefits and compare the treatments' lower bounds. In a similar vein, when comparing two treatments' harms, a patient might be primarily interested in avoiding the highest possible rate of harms and compare the treatments' upper bounds. This strategy is also described in the minimax heuristic according to which people only compare the worst possible outcome and choose the one with the more attractive worst outcome (Savage, 1951). This strategy has been found to predict people's affect-rich choices (Pachur, Hertwig, & Wolkewitz, 2012). One can also shift attention to the option with the best possible outcome (upper bound in benefits condition, lower bound in harms condition). Finally, the two average rates—indicated by the midpoint of the range in the ambiguous option—can serve as basis for comparison. In the benefits condition, a higher average rate signals a better treatment, whereas in the harms a lower average rate does.

In Study 2, we focused on the question which choice strategy underlies people's treatment choice. Based on the findings from Experiments 1a and 1b, we expect rather high variability in people's choices. For instance, when only the degree of ambiguity differed, half of participants opted for the certain and the other half for the ambiguous option. Only few studies investigated people's choice strategies depending on numeracy. For instance, Pachur

and Galesic (2012) found that high numerates often followed the minimax heuristic in risky choice, whereas low numerates were more affective. However, how low and high numerates choose under ambiguity when benefits or harms has not yet been investigated.

Method

Design and Procedure. The procedure of Experiment 2 was similar to Experiment 1a/b, except that participants were randomly assigned to only one out of six hypothetical medical scenarios.[4] In set I, both options had the same average rates, but differed in their degree of ambiguity; in set II and III also the average rates differed (Table 3).

Table 3.

The Six Scenarios of Experiment 2. Random Halves of Participants Were Told that the Treatment Information Is Either About Benefits or Harms.

Set	Ambiguity	Scenario	Treatment 1	Treatment 2
I	certain vs. high	A	20	10 – 30
(same average rates, low vs. high ambiguity)	low vs. high	B	18 – 22	10 – 30
II	low vs. certain	C	18 – 22	17
(different average rates, low ambiguity)	low vs. high	D	18 – 22	7 – 27
I	high vs. certain	E	10 - 30	17
(different average rates, high ambiguity)	high vs. low	F	10 – 30	15 – 19

[4] From Experiment 1a, we selected one scenario from each of the three sets; from Experiment 1b, we took the scenario from set I, and both from set III. Thereby, we included scenarios with either the same or different average rates and included scenarios that showed low and high variance in treatment choice in Experiment 1a/b

We assessed participants' respective choice strategies by collecting verbal protocols. Each participant had to briefly state the major reason for choosing a particular option. Numeracy was again included with the Berlin Numeracy Test.

Two independent raters coded participants' choice strategies. Interrater-reliability was high (Cohen's Kappa = .8). Each participant's written statement was coded according to one of the four categories mentioned below. A fifth category was added for answers which did not match any choice strategy.

(1) *Ambiguity aversion*: The participant states that she selected the treatment with the smaller range (i.e., the one which is less ambiguous). This strategy corresponds to the notion of ambiguity aversion. Example: "The [treatment 1] offers an exact number of people who appeared symptom-free after three days; whereas the [treatment 2] had numbers ranging from 10 – 30 making it appear that they were unsure of how many people were actually symptom- free."

(2) *Midpoint*: The participant states that she compared the average rate of the two treatments, independently of the range. Example: "Because 17 is actually a little lower than 20 which is the midpoint between 10 and 30."

(3) *Lower Bound*: The participant states that she compared both treatments' lower bound (which equals the average rate if no range is present): "The minimum amount of people benefiting from the treatment is higher."

(4) *Upper Bound*: The participant states that she compared both treatments' upper bounds (which equals the average rate if no range is present): Example: "Up to 30 people were symptom-free vs. 17"

(5) *Unidentified*: Some participants' answers could not be identified. Example: Most participants in scenario C simply justified the selection of treatment 1 with "better statistics". Hence, this scenario resulted in a high number of unidentified answers.

Note that the choice strategies have different implications for the benefits and harms condition. In the benefits condition, a higher midpoint signals a better treatment; a focus on the lower bound signals that participants focused on the treatments worst possible outcomes (the lowest possible rate of benefits); a focus on the upper bound signals that participants focused on the treatments with the best possible outcome (the highest possible rate benefits). For harms, the lower the midpoint the better a treatment; the lower bound here signals the best possible outcome (lowest possible rate of harms); the upper bound the worst possible outcome (highest possible rate of harms). If ambiguity averse, one should choose the treatment with the smaller range, independent of whether benefits or harms were at stake.

Participants. In total, 514 participants took part in this Experiment (n=261 in the benefits condition and n=253 in the harms condition) of which 45% were male. The average age was 32.6 *(SD*=12.9). Participants were recruited using Amazon MTurk. Each participant was randomly assigned to one of the six scenarios. Remuneration was $0.75.

Results

Treatment Choice and Reasoning. Participants' treatment choices in Study 2 were consistent with the findings of Experiment 1a/b (see Appendix). In the following, we will present results for each scenario's choice and participants choice strategies. First, we show findings for the benefits condition (Table 4).

In set I—same average rate, low versus high ambiguity—60% of the participants chose the certain option (two-tailed binomial, p=*.203*) in scenario A (see Appendix). Were participants who chose the certain option ambiguity averse or decided based on the lower bound? Two thirds of those followed an ambiguity aversion strategy whereas 20% made choices based on the lower of the two lower bounds of the range (i.e., they avoided the lowest possible rate of benefits). Of the participants who made choices in favor of the ambiguous

option, 95% made their choice based on the higher of the upper bounds of the two treatments (i.e., they focused on the best possible rate of benefits). In scenario B—different average rate, low ambiguity—half of the participants chose the less ambiguous option (51%; two-tailed binomial, *p=1*). Of those, 26% stated that they chose this option due to its smaller range (i.e., ambiguity aversion), whereas 57% made choices based on the worst possible outcome (i.e., lower bound). Of all participants who chose the highly ambiguous option, 96% made their choices due to the higher upper bound (i.e., higher possible rate of benefits).

In set II—different average rate, high ambiguity—a clear majority of participants made choices in favor of treatment 1 in both scenarios (91% and 93% for scenario C and D, respectively, see Appendix). As mentioned above, it was not possible to disentangle participants' choice strategies in scenario C. In scenario D, half of the participants (55%) who chose treatment 1 made choices based on the midpoint (which was higher than the average rate of treatment 2); 30% indicated that they were ambiguity averse (choices in favor of treatment 1 due to its smaller range), and 15% made choices based on the higher of the two lower bounds. All participants who chose treatment 2 made choices based on the higher of the two upper bounds.

In line with Experiment 1a/b, set III yielded higher variance in participant's choices (see Appendix). About two thirds of participants chose the highly ambiguous treatment 1 in scenario E and F (two-tailed binomial, *p=.02*; *p=.038*). In scenario E, 69% chose treatment 1 based on the midpoint and 31% based on the upper bound; when making a choice in favor of the less ambiguous treatment (treatment 2), 39% indicated that they were ambiguity averse and 23% made choices based on the lower bound. In scenario F, 44% chose treatment 1 based on the average rate and 48% based on the upper bound. Those who chose the less ambiguous treatment 2 indicated to be ambiguity avers (23%) or made choices based on the lower bounds (54%).

Table 4

Results of the Open Question in the Benefits Condition (Absolute Number in Parenthesis).

Scenario	Treatment 1					Treatment 2				
	Ambiguity aversion	Midpoint	Lower Bound	Upper Bound	Unidentified	Ambiguity aversion	Midpoint	Lower Bound	Upper Bound	Unidentified
A (20 vs. 10-30)	67% (20/30)	7% (2/30)	20% (6/30)	3.3% (1/30)	3.3% (1/30)	-	-	-	95% (19/20)	5% (1/20)
B (18-22 vs. 10-30)	26% (6/23)	-	57% (13/23)	-	17% (4/23)	-	-	-	96% (21/22)	4% (1/22)
C (18-22 vs. 17)	-	5% (2/41)	2% (1/41)	2% (1/41)	91% (39/41)	-	33% (1/3)	-	-	67% (2/3)
D (18-22 vs. 7-27)	30% (12/40)	55% (22/40)	15% (6/40)	-	-	-	-	-	100% (3/3)	-
E (10-30 vs. 17)	-	69% (20/29)	-	31% (9/29)	-	39% (5/13)	15% (2/13)	23% (3/13)	-	23% (3/13)
F (10-30 vs. 15-19)	-	44% (12/27)	-	48% (13/27)	7% (2/27)	23% (3/13)	-	54% (7/13)	-	23% (3/13)

Which strategies underlay participants' choices when harms were presented (Table 5)? In scenario A of set I (see Appendix), 54% chose the certain option (two-tailed binomial, $p=.85$. Participants indicated ambiguity aversion (47%) and the higher of the two upper bounds (27%) as choice strategy. Participants who chose the high ambiguity option indicated the lower bound (the best possible outcome) as choice strategy (85%). In scenario B, half of the participants chose the less ambiguous option (two-tailed binomial, $p=.39$) and indicated either ambiguity aversion (32%) or the higher of the two upper bounds (46%) as choice strategy. The other half chose the option that had high ambiguity indicating the lower bound (best possible rate of harms) as choice strategy (86%).

In set II, 98% chose treatment 2 in scenario C—as mentioned before, it was not possible to indicate their choice strategy. In scenario D, 92% chose treatment 2 (see Appendix)—41% indicated the midpoint and 50% the lower of the two lower bounds as choice strategy. Participants choosing treatment 1 indicated primarily the upper bound as choice strategy.

In set III, there was a slightly higher variance in choices than in set II. In scenario E, 83% of participants chose treatment 2 (two-tailed binomial, $p=.001$)—the majority (58%) indicated the midpoint, one third indicated ambiguity aversion and 6% the lower of the two upper bounds as choice strategy. When participants chose treatment 1 (86%), they indicated the lower of the two lower bounds as choice strategy. In scenario F, 79% of participants chose treatment 2 (two-tailed binomial, $p=.001$)—the majority 49% chose this option due to the midpoint, 24% the lower of the two upper bounds and 17% ambiguity aversion. Again, when participants chose treatment 1, they indicated the lower of the two lower bounds (46%) as choice strategy.

Table 5

Results of the Open Question in the Benefits Condition (Absolute Number in Parenthesis).

	Treatment 1					Treatment 2				
Scenario	Ambiguity aversion	Midpoint	Lower Bound	Upper Bound	Unidentified	Ambiguity aversion	Midpoint	Lower Bound	Upper Bound	Unidentified
A (20 vs. 10-30)	47% (7/15)	13% (2/15)	7% (1/15)	27% (4/15)	7% (1/15)	-	-	85% (11/13)	8% (1/13)	8% (1/13)
B (18-22 vs. 10-30)	32% (9/28)	-	4% (1/28)	46% (13/28)	18% (5/28)	-	-	86% (18/21)	-	14% (3/21)
C (18-22 vs. 17)	-	-	-	-	100% (1/1)	2% (1/48)	-	-	-	98% (47/48)
D (18-22 vs. 7-27)	67% (2/3)	-	-	-	33% (1/3)	-	41% (14/34)	50% (17/34)	-	9% (3/34)
E (10-30 vs. 17)	-	-	86% (6/7)	-	14% (1/7)	30% (10/33)	58% (19/33)	3% (1/33)	6% (2/33)	3% (1/33)
F (10-30 vs. 15-19)	-	-	46% (5/11)	-	54% (6/11)	17% (7/41)	49% (20/41)	2% (1/41)	24% (10/41)	7% (3/41)

Numeracy. We categorized participants into low and high numerates (low: 34%, high: 65%, missing: 1%). In line with Experiment1a/b, low numerates showed higher variance in their choices than high numerates. Particularly in set III in the harms condition, 29% and 35% of low numerates compared with 11% and 16% of high numerates made choices in favor of treatment 1 (see Appendix). In other words, more low numerates aimed for the best possible outcome of the harms (focusing on the lower bounds as choice strategy). When it comes to the underlying choice strategies, we will only discuss scenarios, in which low and high numerates' choice strategies differed, for reasons of brevity. It should be noted that due to a limited sample size of low numerates, results are only tentative.

We first present differences for the benefits condition: In set I, more high numerates than low numerates in scenario A and B indicated ambiguity aversion as choice strategy to choose the certain option (scenario A: 78% of high vs. 28% of low numerates; scenario B: 38% of high vs. 10% of low numerates). In other words, a higher proportion of low numerates than high numerates compared the lower bounds of the two treatments (scenario A: 28% of low vs. 8% of high numerates; scenario B: 85% of low vs. 44% of high numerates). In set III, 86% of low numerates compared with 53% of high numerates indicated the midpoint and 14% of low numerates compared with 47% of high numerates indicated the upper bound as choice strategy. There were no differences in high and low numerates for those choosing treatment 2.

When it comes to harms, we again found a higher proportion of high numerates choosing the certain indicating ambiguity aversion as choice strategy (scenario A: 20% of low vs. 60% of high numerates; scenario B: 13% vs. 40%); however, more low numerates who chose the certain option indicated the lower of the two upper bounds as choice strategy (scenario A: 30% of low vs. 20% of high numerates; Scenario B: 62% of low vs. 40% of high

numerates). In set II and III, we did not find any differences in choice strategy due to the higher variance.

Discussion Study 2

Findings of Experiment 2 were in line with Experiment 1a/b—participants made similar choices within the scenarios and the variance in choices was higher in low numerates than in high numerates. Moreover, we mapped participants' choice strategies and thereby extended previous findings. Overall, we observed a large heterogeneity in participants' choice strategies; there was no dominant choice strategy. Even when a majority of participants made choices in favor of one treatment, the underlying strategies were quite diverse. Interestingly, when both options only differed in the degree of ambiguity, about half of the subjects were optimistic. More precisely, in the benefits condition, half focused on the upper bound of the ambiguous option, and in the harms condition, half focused on the lower bound of the ambiguous option.

Differences between low and high numerates indicate that a higher proportion of low numerates focused on the lower of the two lower bounds in the benefits condition and higher of the two upper bounds in the harms condition. However, when low and high numerates chose the same option differences in choices strategies were small.

In sum, whereas research suggests that ambiguity makes people uncomfortable due to lack of knowledge (Fox & Tversky, 1995) or the shift in attention to the worst possible outcome (Viscusi et al., 1991), the present results indicate that individuals deal differently with ambiguity. Numeracy is one potential moderator. We found similar differences in choice patterns between low and high numerates, but when choosing the same option, the underlying choice strategies were similar between both groups. Further dispositional factors might help to explain people's different choice strategies. Yet, situational factors—the degree of

ambiguity and differences in the average rates—play an important role and trigger different choice strategies. For instance, more people stated ambiguity aversion as their choice strategy when one option was certain and one ambiguous. Presenting both option as ambiguous reduced ambiguity aversion, and more people shifted attention to the respective bounds.

General Discussion

The present findings may question the general claim that people cannot choose medical treatments under ambiguity. In fact, the vast majority of the participants in our studies chose the "superior" treatment in most scenarios. Whether ambiguity increases complexity in the medical choice behavior rather depends on the distinctiveness of the treatment outcomes—the more choice strategies favor a specific option, the smaller variance in choices should be. For instance, when the range of options does not overlap, ambiguity is very informative as one option can be clearly identified as superior. When the range of options overlaps, complexity is high as one or more strategies make contradicting predictions. Yet, we found a majority of participants choosing the same option. Those who deviated from majority choice were either more concerned about the worst possible outcome or best possible outcome.

Our findings have at least two important theoretical implications. First, there was no evidence for ambiguity aversion in treatment choice when benefits were presented. Moreover, we did not find evidence for ambiguity seeking when harms were presented. This challenges classical views on people's behavior under ambiguity. One potential explanation is the context in which decisions take place. Previous research on ambiguity aversion and seeking was based on fictitious monetary gambles. In such situations an optimal solution is mathematically traceable—a standard based on which behavior is evaluated. However, monetary gambles hardly reflect real-world decisions. The medical scenarios represent an important context in peoples' lives. In this context, causal mechanisms have a long history

and play an important role. For instance, throughout their evolutionary history, people had to quickly learn whether a specific substance causes an illness or death. Therefore, medical information might be encoded differently and is not easily changed by numerical evidence only (Müller, Garcia-Retamero, Galesic, & Maldonado, 2012).

Second, in contrast with previous research on ambiguity aversion, we found a diversity of choice strategies. This diversity depends on at least two factors: the scenario (situational factors) and individual differences (dispositional factors). Depending on the treatments' effectiveness and the degree of ambiguity, participants applied different choice strategies. This indicates that concepts like ambiguity aversion or seeking are not universal, but have to be evaluated in light of factors such as the context (medical, financial), the frame (gain vs. loss), the perspective (choosing for oneself vs. choosing for others), and the given odds and degree of ambiguity. The heterogeneity in choice strategies could also be partially a consequence of individual differences. Some participants' choices represented a rather optimistic view on the expected outcomes; others' choices represented a pessimistic and more conservative strategy. Han and colleagues (2010) also argue that dispositional factors such as optimism may account for individual differences in choice strategies. Numeracy seems to moderate the results only slightly. In line with previous findings, low numerates had little more difficulties in identifying a superior option when it existed (e.g., Peters et al., 2006; Pachur & Galesic, 2012). Yet, we found that low and high numerates had similar choice strategies, although more low numerates were concerned about the lower bounds in the benefits (pessimistic treatment choice) as well as in the harms condition (optimistic treatment choice).

These findings have important practical implications. In many medical decisions, it is not possible to define an optimal solution (Feufel & Bodemer, 2012). In these situations, it is important to communicate complete information, so people can decide based upon their

individual preferences, values, and needs. These differences may be presented by the heterogeneity in choice strategies in our studies. An interesting example is the implementation of the human papilloma virus (HPV) vaccine. In Germany, researchers questioned the admission of the vaccine due to a lack of knowledge about its effectiveness and potential side-effects (Dören et al., 2008). However, statistical information and in particular information about the ambiguity of effectiveness and side-effect estimates were hardly provided in patient information (Bodemer et al., 2012). Yet, due to the controversy and limited evidence, such information would have been important for a decision maker—primarily parents and young girls—to evaluate the level of evidence and decide whether to get vaccinated or not.

The present findings also point to future lines of research. First, only little is known about how adaptive people are to situational factors when ambiguity is present. For instance, Leonhardt and colleagues (2011) found that when choosing for others, people actually tend to seek for ambiguity, but not when choosing for themselves. Moreover, different domains may trigger different choice strategies (Müller et al., 2012). People's tendency to be risk seeking or shift attention to the best possible outcome may also depend on the severity of a disease, or how effective a treatment actually is. For instance, in our examples the benefits were rather low (on average 20 out of 100). Second, individual differences to better tailor information to patients' require a better understanding not only of situational factors, but also on dispositional factors such as optimism or tolerance for uncertainty (Furnham & Ribchester, 1995). Third, we simplified treatment choice by keeping either the benefits or harms constant. In most medical decisions, both have to be taken into consideration and may point into different choices. For instance, a treatment with a higher benefit may go along with a higher risk of harms. This adds complexity and potentially influences choice strategies. A fourth line of research points to the presentation format of ambiguity. In line with previous research, we presented ambiguity with a range (e.g., Han et al., 2009, 2010; Schapira et al., 2001).

However, little is known about how people interpret a range and how it affects choice strategies. The fact that laypeople are rather unfamiliar with the concept of ambiguity and its presentation (Ibrekk & Morgan, 1988; Politi et al., 2007), raises the question how people actually interpret such information. For instance, whereas one might perceive the midpoint of the range as the most likely value, one could also assume that each value within the range is equally likely. We need to better understand how people—in particular low numerates—interpret the range, and whether we can present ambiguity more intuitively. For instance, graphical tools such as icon arrays (pictographs) have been shown to improve people's understanding of risks (Bosnjak & Pahl, 2011; Galesic, Garcia-Retamero, & Gigerenzer, 2009; Gaissmaier et al., 2011)

In sum, the communication of ambiguity in medical risk communication increases transparency and helps patients to choose treatments in line with their individual preferences. We argue that participants can and should be informed about the ambiguity of treatment benefits and harms to satisfy their individual preferences, needs, and values.

Author Note

We thank Azzurra Ruggeri, Mirta Galesic and Henrik Olsson for helpful comments on former versions of this manuscript.

References

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association*, *13*(6), 608–618. doi:10.1197/jamia.M2115

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision making. *Judgment and Decision Making, 6*, 252-262.

Bodemer, N., Müller, S. M., Okan, Y., Garcia-Retamero, R., & Neumeyer-Gromen, A. (2012). Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine. *Vaccine*, *30*(25), 3747–3756. doi:10.1016/j.vaccine.2012.03.005

Bosnjak, M. & Pahl, H. (2011). *Risikokommunikation in Medizinischem Marketing mit Laien als Zielgruppe: Effekte der Berücksichtigung von Unsicherheiten bei der Kommunikation von Risiken*. Working Paper, University of Mannheim.

Braddock, C. H., Edwards, K. A., Hasenberg, N. M., Laidley, T. L., & Levinson, W. (1999). Informed decision making in outpatient practice: time to get back to basics. *The Journal of the American Medical Association*, *282*(24), 2313–2320.

Briss, P., Rimer, B., Reilley, B., Coates, R. C., Lee, N. C., Mullen, P., Corso, P., et al. (2004). Promoting informed decisions about cancer screening in communities and healthcare systems. *American Journal of Preventive Medicine*, *26*(1), 67–80.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. doi:10.1007/BF00122575

Curley, S. P., & Yates, J. F. (1985). The center and range of the probability interval as factors affecting ambiguity preferences. *Organizational Behavior and Human Decision Processes*, *36*(2), 273–287. doi:10.1016/0749-5978(85)90016-0

Dören, M., Gerhardus, A., Gerlach, F.M., Hornberg, C., Kochen, N.M., Kolip, P., et al. (2008). *Wissenschaftlier/innen fordern Neubewertung der HPV-Impfung und ein Ende der irreführenden Informationen* [Internet]. 2008 Nov [cited 2011 Feb 1]. Retrieved from: http://www.uni-bielefeld.de/gesundhw/ag3/downloads/Stellungnahme_Wirksamkeit_HPV-Impfung.pdf.

Edwards, A., & Elwyn, G. (2009). *Shared decision-making in health care: Achieving evidence-based patient choice*. Oxford; New York: Oxford University Press.

Einhorn H. J. & Hogarth, R. M. (1986). Decision making under ambiguity. *The Journal of Business*, 59, 225–50.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, *75*(4), 643. doi:10.2307/1884324

Epstein, L. G. (1999). A Definition of Uncertainty Aversion. *Review of Economic Studies*, *66*(3), 579–608. doi:10.1111/1467-937X.00099

Farrell, M. H., Murphy, M. A., & Schneider, C. E. (2002). How underlying patient beliefs can affect physician-patient communication about prostate-specific antigen testing. *Effective Clinical Practice: ECP*, *5*(3), 120–129.

Feufel, M. A., Antes, G., & Gigerenzer, G. (2010). Vom sicheren Umgang mit Unsicherheit: Was wir von der pandemischen Influenza (H1N1) 2009 lernen können. *Bundesgesundheitsblatt*, *53*(12), 1283–1289. doi:10.1007/s00103-010-1165-1

Feufel, M.A. & Bodemer, N. (2012). *Finding the right tool to improve health decisions: Nudging, social marketing, empowerment*. Manuscript in preparation.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, *110*(3), 585–603. doi:10.2307/2946693

Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research*, *6*(1), 75–85. doi:10.1080/1366987032000047815

Frosch, D. L., Kaplan, R. M., & Felitti, V. (2001). Evaluation of two methods to facilitate shared decision making for men considering the prostate-specific antigen test. *Journal of General Internal Medicine*, *16*(6), 391–398. doi:10.1046/j.1525-1497.2001.016006391.x

Furnham, A., & Ribchester, T. (1995). Tolerance of ambiguity: A review of the concept, its measurement and applications. *Current Psychology*, *14*(3), 179–199. doi:10.1007/BF02686907

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210–216. doi:10.1037/a0014474

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: overcoming challenges in people's understanding of risks. *Social Science & Medicine (1982)*, *70*(7), 1019–1025. doi:10.1016/j.socscimed.2009.11.031

Gigerenzer, G. (2003). *Reckoning with risk: Learning to live with uncertainty*. London: Penguin.

Gigerenzer, G. (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge: MIT Press.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. doi:10.1111/j.1539-6053.2008.00033.x

Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2011). Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*. doi:10.1177/0272989X10393976

Han, P. K. J., Klein, W. M. P., Lehman, T. C., Massett, H., Lee, S. C., & Freedman, A. N. (2009). Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. *Medical Decision Making*, *29*(3), 391–403. doi:10.1177/0272989X08327396

Han, P. K. J., Klein, W. M. P., Lehman, T., Killam, B., Massett, H., & Freedman, A. N. (2010). Communication of uncertainty regarding individualized cancer risk estimates: effects and influential factors. *Medical Decision Making*, *31*(2), 354–366. doi:10.1177/0272989X10371830

Heath, C. & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty, 4*, 5-28.

Higgins, J. P. T., Green, S., & Cochrane Collaboration. (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, England; Hoboken, NJ: Wiley-Blackwell.

Ibrekk, H., & Morgan, M. G. (1987). Graphical Communication of Uncertain Quantities to Nontechnical People. *Risk Analysis*, *7*(4), 519–529. doi:10.1111/j.1539-6924.1987.tb00488.x

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk Analysis*, *15*(4), 485–494.

Keren, G., & Gerritsen, L. E. M. (1999). On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, *103*(1-2), 149–172. doi:10.1016/S0001-6918(99)00034-7

Laughhunn, D. J., Payne, J. W., & Crum, R. (1980). Managerial risk preferences for below-target returns. *Management Science*, *26*(12), 1238–1249. doi:10.1287/mnsc.26.12.1238

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44.

Müller, S. M., Garcia-Retamero, R., Galesic, M., & Maldonado A. (2012).*The impact of domain specific beliefs on decisions and causal judgments*. Manuscript under review.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407–413. doi:10.1111/j.1467-9280.2006.01720.x

Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making, 3*, 435 – 448.

Politi, M. C., Han, P. K. J., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695. doi:10.1177/0272989X07307270

Pachur, T., & Galesic, M. (2012). Strategy selection in risky choice: The impact of numeracy, affect, and cross-cultural differences. *Journal of Behavioral Decision Making*, doi:10.1002/bdm.1757

Pachur, T., Hertwig, R., & Wolkewitz, R. (2012). *Money, pleasure, or pain: Reversals in risky choice between affective and nonaffective domains*. Manuscript in preparation.

Raffle, A. E. (2001). Information about screening - is it to achieve high uptake or to ensure informed choice? *Health Expectations*, *4*(2), 92–98.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. doi:10.1037/a0017327

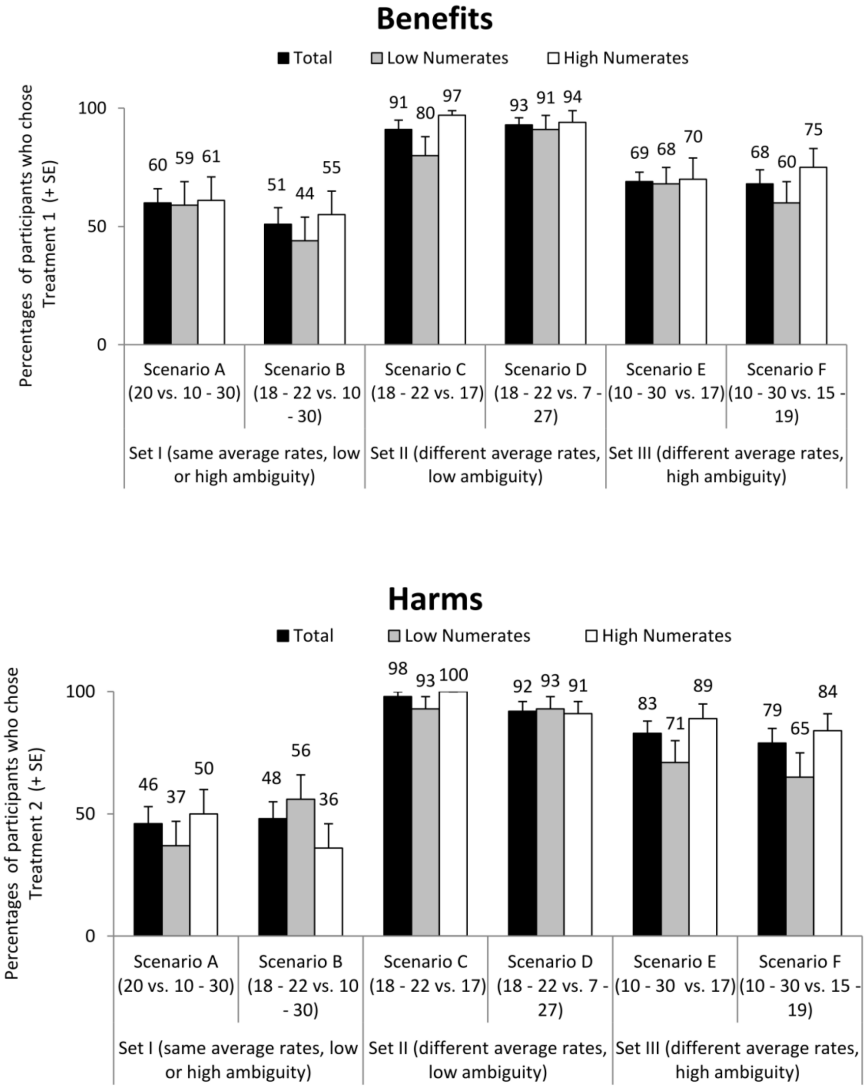
Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association, 46*, 55 – 67.

Schapira, M. M., Nattinger, A. B., & McHorney, C. A. (2001). Frequency or Probability? A Qualitative Study of Risk Communication Formats Used in Health Care. *Medical Decision Making*, *21*(6), 459–467. doi:10.1177/0272989X0102100604

Taplin, S. H., Urban, N., Taylor, V. M., & Savarino, J. (1997). Conflicting national recommendations and the use of screening mammography: does the physician's recommendation matter? *The Journal of the American Board of Family Practice*, *10*(2), 88–95.

Visucsi, W.K., Magat, W.A.,& Huber, J. (1991) Communication of ambiguous risk information. *Theory and Decision*, 31, 159-173.

Appendix



Percentages of participants who chose treatment 1 in Experiment 2 – benefits condition (above: total *n* = 261, low numerates n = 87, high numerates *n* = 170) and percentages of participants who chose treatment 2 in Experiment 2 – harms condition (below: total *n* = 253, low numerates n = 81, high numerates *n* = 171).

Chapter 6

General Discussion

General Discussion

In this dissertation, I presented theoretical and empirical research that contributes to a better understanding of how we can help patients making informed decisions. Using theoretical concepts and methods from psychology to gain insights into the mechanism of how laypeople understand and perceive risks, I derived important implications for future research and practice.

Building a basis for a health care system in which patients actively and responsibly participate in their medical decisions requires transparent, complete, and intuitive information. Therefore, we have to be aware of (1) which tools are available to foster "better" health care decisions, (2) the shortcomings in current risk communication, (3) how biased formats undermine the empowerment of patients, and (4) how we can present complete information including ambiguity. In this final chapter, I will briefly summarize what we have learned from the papers composing this dissertation and give an outlook of possible future directions for each of them.

(1) Empowering patients—a matter of perspective

There are two perspectives about the human rationality. Some argue that human's use heuristics that are prone to biases leading to poor decisions (Kahneman & Tversky, 1983). These biases have been primarily documented in what Savage (1954) calls "small worlds"—worlds in which an optimal solution can be computed. In other words, such situations are characterized by perfect knowledge about alternatives and probabilities. Bias is defined as human's deviation from such a normative standard, being taken as the proof that people are not rational. As a consequence, people need guidance in form of nudges to overcome these biases (Marteau, Ogilvie, Roland, Suhrcke, & Kelly, 2011; Thaler & Sunstein, 2009). Nudging is based on the idea to design environments that prompt a particular behavior

without restricting any options. Yet, this perspective about the human mind is rather pessimistic. In most situations in our daily life, rarely—if ever—all alternatives and probabilities are known. Decisions are to be made under considerable uncertainty due to limited knowledge and limited capacities. In such situations an optimal solution is intractable and heuristics have been found to be very effective strategies (Gigerenzer, Hertwig, & Pachur, 2011; Todd, Gigerenzer, & the ABC Research Group, 2012). The term ecological rationality has been coined to describe the structure and representation of information in an environment and the match with mental strategies. Hence, the problem is less in the human mind, but in the way information is presented. Intuitive design follows this principle and aims at designing environments that match cognitive processes. One prominent example is the presentation of statistical information to patients in form of natural frequencies in contrast to conditional probabilities (Gigerenzer & Hoffrage, 1995).

These two perspectives about the human mind have led to different approaches to improve health decisions: on the one hand, nudging and social marketing resting on the former assumption, on the other hand empowerment resting on the latter assumption. In Chapter 2, I presented the differences, commonalities and applicability of these three approaches. Nudging and social marketing are limited to situations in which a normative standard can be clearly defined, for instance, when there is strong evidence proposing an "optimal" solution. However, in many medical situations patients and health professionals have to make decisions under uncertainty—an "optimal" solution does not exist. Moreover, what is "good" or "bad" is often not ultimate, but depends on the patient, her values and needs. Therefore, instead of imposing one solution, empowerment aims at transparently informing and educating patients to make medical decisions that suit their personal situation best. This builds a basis for shared decision making and informed consent (Gigerenzer & Gray, 2011).

However, the implementation of empowerment has to meet at least three challenges: First, like the nudging and social marketing approaches, positive effects of empowerment need to be proven. Therefore, empowerment strategies have to be evaluated on relevant health outcomes. Can empowerment reduce inequality in health care practice such as regional variability? Can it increase quality of life in patients? Can it make the health care system more cost-efficient? The second challenge addresses how empowerment can be implemented. Conflicts of interest in politics, industries and health professionals undermine efforts to educate patients (Gigerenzer & Gray, 2011). Patients have to learn which questions to ask and where to find transparent information. Independent institutions are one step towards providing transparent and complete information and help to educate future, risk literate generations. Third, to design environments based on the principle of ecological rationality, we need to better understand how patients actually process statistical information and make medical decisions. Which information is relevant for patients to decide between treatment alternatives? How can we communicate this information transparently to different target groups? A starting point to examine patient decision strategies and heuristics is given by the concepts of bounded and ecological rationality (Gigerenzer, Todd, & the ABC Research Group, 1999).

(2) A tool for empowerment: The media

The media are one channel to empower patients (Grilli, Ramsay, & Minozzi, 2009). But are media reports based on transparent and complete information? In the third chapter, I presented the role of the media in medical risk communication, its opportunities and shortcomings. Based on recent findings in risk communication, I developed a coding scheme to evaluate reporting about the HPV vaccine in German and Spanish newspaper and Internet reports. The results showed that media reports hardly met the standards proposed for transparent, complete and correct information. Although the Internet reports communicated

more relevant information about the HPV vaccine, they still did not provide a basis for making informed decisions.

Evaluating how the environment presents information is one part in the study of ecological rationality. The second part focuses on how people actually perceive and process the information provided and which consequences it has on behavioral outcomes. Few studies have demonstrated the influence of the media, for instance media reporting on health service utilization (Grilli et al., 2009). To shed more light on how the media—and particularly the Internet—shapes risk perception and decision making, we need to gain insights about how patients actually seek for information, which strategies they use and how they evaluate information and its communicators (Eysenbach & Köhler, 2002; Eysenbach, Powell, Kuss, & Sa, 2002; Feufel & Stahl, 2012; Hesse, Nelson, Kreps, Croyle, Arora, Rimer, & Viswanath, 2005). For instance, Feufel and Stahl (2012) found that web users often stop web search after the first piece of evidence satisfying search intentions is found. Differences in age and web-use skills also moderated participants' search strategies. How such differences in strategies are influenced by the way the media present information can serve as a basis for changing the practice of media coverage about health issues as well as designing interventions for patients searching for such information.

(3) Including baseline risk when communicating relative risk reductions

One prominent example of biased reporting are relative risk reductions and increases. Laypeople as well as health professionals overestimate benefits or harms when changes in risk are expressed in relative rather than absolute terms (Akl et al., 2011; Edward, Elwyn, Covey, Matthews, & Pill 2001). Yet, relative risk changes are still predominant in risk communication (Gigerenzer, Wegwarth, & Feufel, 2010; Sedrakyan & Shih, 2007). Their use is sometimes defended with an argument that relative risk changes can be "debiased" if

baseline risk is included (Natter & Berry, 2005; Schwartz, Woloshin, Black, & Welch, 1997). In four experiments, I tested the influence of the presentation format (percentage vs. frequency) and people's numeracy abilities on the interpretation of relative risk changes with baseline risk. Results showed that the understanding of relative risk changes with baseline risk depends on (i) the presentation format used to communicate the baseline risk (percentage vs. frequencies) and (ii) people's numeracy skills. Whereas high numerates understood relative risk reductions and increases independently of the presentation format, low numerates benefited only when the baseline risk was presented in frequencies rather than in percentages. Yet, we found that—independently of the presentation format and numeracy abilities—many participants still failed to correctly understand the information. Thus, relative risk changes, even when communicated with baseline risk, remain a source of confusion. This questions whether these formats are suitable in practice.

Future research can further investigate the role of the presentation format on people's understanding of risk changes in dependence of numeracy. One stream of research tries to avoid numerical information as far as possible, and instead displays risks visually. For instance, icon arrays (pictographs) can help to overcome low numeracy (Galesic, Garcia-Retamero, & Gigerenzer, 2009; Garcia-Retamero & Galesic, 2010). However, only for those high in graph literacy visualization might be better than numerical formats, whereas for people low in graph literacy the opposite may be true (Gaissmaier, Wegwarth, Skopec, Müller, Broschinski, & Politi, 2011). Most statistics—may it be numbers or graphs—are so called descriptive statistics; they summarize statistical evidence. However, recent research in risky choice has shown that decisions about which option to choose may depend on whether one is presented with descriptive statistics or one actively samples outcomes separately (Hertwig, Barron, Weber, & Erev, 2004; Ungemach, Chater, & Stewart, 2009). These

findings also have implications for risk communication and influence people's treatment choice (Bodemer, Gaissmaier, & Nelson, 2012).

Whereas most research focuses on one presentation format only, another issue is how laypeople and experts integrate and compare treatment effects that are framed in different formats. The phenomenon of mismatched framing—benefits are presented in relative risk reductions to appear large, whereas harms are presented in absolute risk increases to appear small—amplifies confusion (Gigerenzer et al., 2010). To make sense of statistics framed in different formats, one has to find a common denominator to compare the magnitude of effects. Frequency formats help people reasoning about statistical concepts (e.g., Moster, 2002), but future research may address whether people transform percentage formats into frequency formats, or vice versa.

(4) Disclosing uncertainty: Presenting ambiguity in risk communication

Providing complete information also includes the communication of uncertainty. Medical evidence is often limited, but this is rarely presented to patients (Bodemer, Müller, Okan, Garcia-Retamero, & Neumeyer-Gromen, 2012; Politi, Col, & Han, 2007). In Chapter 5, I questioned general objections against ambiguity communication such as people's inability to handle it (Frewer, Hunt, Brennan, Kuznesof, Ness, & Ritson, 2003). Results showed that participants were able to find a superior treatment option even when ambiguity was presented. Moreover, heterogeneity in participants' choice strategies extends previous research and demonstrates how important it is to consider individual differences in treatment choice under ambiguity. For instance, some participants were pessimistic and chose based on the worst possible outcome, others were optimistic and chose based on the best possible outcome. We suggest including ambiguity when presenting treatment benefits and harms to provide complete information about potential limitations of the existing evidence.

As mentioned in the previous section, presenting statistics in numerical format are only one way to communicate risks. Although I found only limited evidence for differences in numeracy, some studies suggest that numeracy plays a major role in medical decision making (Pachur & Galesic, 2012; Reyna, 2009). Hence, graphical formats present a promising alternative to visualize treatment effects with ambiguity. For instance, Bosnjak and Pahl (2011) propose bar charts with a confidence interval as an intuitive format to communicate ambiguity without increasing complexity.

Moreover, further situational and dispositional factors are to be explored to provide a more thorough investigation of how people react towards ambiguity. Optimism as well as tolerance for uncertainty might help to predict people's choice under ambiguity (Han, Klein, Lehman, Massett, Lee, & Freedman, 2010). Ambiguity and—in a broader sense—uncertainty is a major component of our environment and provide crucial information for a decision maker—may it be regarding financial, environmental, medical, or social issues.

Conclusion

In sum, the work presented here gave new insights into (medical) decision making, risk perception and risk communication. The findings show that empowerment is a crucial tool to improve health decisions; the media lack transparent communication of health information which calls for standard to improve reporting; one should be careful when communicating relative risk reductions and increases also along with baseline risk; and that ambiguity is an important element in medical decision making.

References

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., Costiniuk, C., et al. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews (Online)*, (3), CD006776. doi:10.1002/14651858.CD006776.pub2

Bodemer, N., Müller, S. M., Okan, Y., Garcia-Retamero, R., & Neumeyer-Gromen, A. (2012). Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine. *Vaccine*, *30*(25), 3747–3756. doi:10.1016/j.vaccine.2012.03.005

Bodemer, N., Gaissmaier, W., & Nelson, J. (2012). *Descriptive versus experience-based risk communication and the influence of narratives*. Manuscript in preparation.

Bosnjak, M. & Pahl, H. (2011). *Risikokommunikation in Medizinischem Marketing mit Laien als Zielgruppe: Effekte der Berücksichtigung von Unsicherheiten bei der Kommunikation von Risiken*. Working Paper, University of Mannheim.

Edwards, A., Elwyn, G., Covey, J., Matthews, E., & Pill, R. (2001). Presenting risk information--a review of the effects of „framing“ and other manipulations on patient outcomes. *Journal of Health Communication*, *6*(1), 61–82. doi:10.1080/10810730150501413

Eysenbach, G. & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, *324*(7337), 573–577. doi:10.1136/bmj.324.7337.573

Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review. *The Journal of the American Medical Association*, *287*(20), 2691–2700.

Feufel, M. A., & Stahl, S. F. (2012). What do Web-Use Skill Differences Imply for Online Health Information Searches? *Journal of Medical Internet Research*, *14*(3), e87. doi:10.2196/jmir.2051

Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research*, *6*(1), 75–85. doi:10.1080/1366987032000047815

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A.-S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, *31*(3), 286–296. doi:10.1037/a0024850

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210–216. doi:10.1037/a0014474

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. *Social Science & Medicine (1982)*, *70*(7), 1019–1025. doi:10.1016/j.socscimed.2009.11.031

Gigerenzer, G., & Gray, M. (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge: MIT Press.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford; New York: Oxford University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (2001). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gigerenzer, G., Wegwarth, O., & Feufel, M. (2010). Misleading communication of risk. *British Medical Journal*, *341*, c4830.

Grilli, R., Ramsay, C., & Minozzi, S. (2009). Mass media interventions: Effects on health services utilisation. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. doi.wiley.com/10.1002/14651858.CD000389

Han, P. K. J., Klein, W. M. P., Lehman, T. C., Massett, H., Lee, S. C., & Freedman, A. N. (2009). Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. *Medical Decision Making*, *29*(3), 391–403. doi:10.1177/0272989X08327396

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., & Viswanath, K. (2005). Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Archives of Internal Medicine*, *165*(22), 2618–2624. doi:10.1001/archinte.165.22.2618

Johnson, E. J. & Goldstein, D. (2003). Do Defaults Save Lives? *Science*, *302*(5649), 1338–1339. doi:10.1126/science.1091721

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge; New York: Cambridge University Press.

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? *British Medical Journal, 342*, d228.

Moser, C. (2002). *Improving Numeracy and Literacy: A Fresh Start*. Nottingham: DFES Publications.

Natter, H. M., & Berry, D. C. (2005). Effects of presenting the baseline risk when communicating absolute and relative risk reductions. *Psychology, Health & Medicine*, *10*(4), 326–334. doi:10.1080/13548500500093407

Pachur, T., & Galesic, M. (2012). Strategy selection in risky choice: The impact of numeracy, affect, and cross-cultural differences. *Journal of Behavioral Decision Making*, doi:10.1002/bdm.1757

Politi, M. C., Han, P. K. J., & Col, N. F. (2007). communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695. doi:10.1177/0272989X07307270

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. doi:10.1037/a0017327

Savage, L. J. (1954). *The foundations of statistics*. New York: Dover Publications.

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

Sedrakyan, A., & Shih, C. (2007). Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. *Medical Care*, *45*(10 Supl 2), S23–28. doi:10.1097/MLR.0b013e3180642f69

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. Oxford; New York: Oxford University Press.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)? *Psychological Science*, *20*(4), 473–479. doi:10.1111/j.1467-9280.2009.02319.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig und ohne die unzulässige Hilfe Dritter verfasst habe und die Dissertation auch in Teilen keine Kopie anderer Arbeiten darstellt. Die verwendeten Hilfsmittel sowie Literatur sind vollständig angegeben. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden. Ich habe keinen Doktorgrad in dem Promotionsfach Psychologie, und die zugrunde liegende Promotionsordnung der Humboldt-Universität vom 03.08.2006 ist mir bekannt.

Berlin, den 20. Juni 2012

_____

Nicolai Bodemer