Patient and Physician Preferences for Reporting Research Findings

Elizabeth Mostofsky, ScD, MPH; Jillian A. Dunn, BA; Sonia Hernández-Díaz, MD, DrPH; Murray A. Mittleman, MD, DrPH

BACKGROUND AND OBJECTIVES: Although absolute risks provide useful information on the number of excess cases attributable to a harmful exposure or the number of fewer cases attributable to an effective treatment, most research findings are reported in terms of multiplicative effects of exposure on the outcome. We aimed to examine patient and physician preferences for communicating research findings in terms of additive and multiplicative measures of effect and measures of heterogeneity of treatment effects.

METHODS: We conducted a cross-sectional study of 76 patients and 26 physicians at an academic medical center in Boston, Massachusetts. Trained study staff conducted interviews about interpretations of study findings.

RESULTS: Physicians were equally likely to prefer the number needed to harm or the risk ratio as the most useful measure for making informed health decisions, and patients strongly preferred risk ratios. Participants perceived changes in risk as larger when they were presented using multiplicative measures such as risk ratios than when presented as additive measures such as the risk differences or number needed to treat or harm.

CONCLUSIONS: Despite the importance of considering absolute risks and benefits of treatment options, patients and physicians who rely on study findings for making informed decisions often prefer relative measures.

Despite a vast literature on the importance of describing disease risk in terms of the absolute increase or decrease in the number of cases among those who are treated compared to those not treated,[1-4] it is common for research findings to be reported in terms of the multiplicative change in risk that is easily calculable with common statistical tools.[5-7] The different implications for informed decision-making of absolute versus multiplicative effects are often ignored. In this study, we aimed to examine patient and physician preferences for reporting research findings.

Assuming no bias or random error, all measures of association provide useful but different information about causal effects. Multiplicative measures such as risk ratios and rate ratios describe how many times greater or lower the risk among the exposed is relative to the unexposed. Additive measures such as risk differences and rate differences describe the absolute number of excess or fewer cases among the exposed relative to the unexposed. Number needed to treat or harm is the inverse of the risk difference that indicates how many patients need to be treated over a specific period to avoid one case or have one additional case.[8-11] People tend to focus on the numerators of a proportion or rate,[12] suggesting that the frequency of events rather than proportions or rates may be most useful.[13] When outcomes are infrequent, changes in risk are perceived as larger when presented as multiplicative measures than when presented as additive measures, and patients and physicians may inappropriately perceive the magnitude of benefit or harm of a given treatment if presented in terms of a relative risk.[13-16] For example, if a treatment increased the risk of an adverse event from 0.1% to 0.2%, the multiplicative measure indicates that the risk in the treated is doubled compared to the untreated. However, the additive

From the Cardiovascular Epidemiology Research Unit, Beth Israel Deaconess Medical Center, Boston, MA (Drs Mostofsky and Mittleman, and Ms Dunn); and Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA (Drs Mostofsky, Hernández-Díaz, and Mittleman).

measure indicates that treatment is associated with a 0.1% increase in the risk of the adverse event, or that 1,000 people would need to receive the treatment to expect to see one excess adverse event. On the other hand, if treatment doubled the risk of the adverse event from 10% to 20%, then only 10 people would need to be treated to lead to one excess adverse event.

Effect measure modification, also known as interaction and heterogeneity of treatment effects, is another source of confusion for patients and physicians. In the presence of heterogeneity of treatment effects, the effect of an exposure on an outcome is different across levels of some third factor. For instance, the effectiveness of a drug on improving a health outcome may depend on a patient's genotype. If an additive measure of association is modified by the third factor, there is additive heterogeneity, and if a multiplicative measure is modified by a third factor, there is multiplicative heterogeneity. Heterogeneity of treatment effects may occur on the additive scale (differences in the absolute risk among people using blood pressure medications), the multiplicative scale (differences in the multiplicative risk among people using blood pressure medications), both scales, or neither scale.[5,17]

Therefore, we aimed to examine patient and physician preferences for communicating research findings in terms of risk ratios, risk differences, number needed to treat or harm, and additive and multiplicative heterogeneity of treatment effects.

Methods

Beth Israel Deaconess Medical Center (BIDMC) is a tertiary care teaching hospital with 673 beds and had 40,656 inpatient discharges in 2017. Embedded within BIDMC is a large, academic primary care practice, Health Care Associates (HCA), with four practice sites across the Boston-based medical center. In 2017, HCA provided care for over 41,500 patients with 271,063 outpatient visits. We identified a convenience sample of patients who receive their primary care at HCA and physicians who provide care for patients hospitalized at the BIDMC.

A convenience sample of HCA primary care providers provided written permission to contact potentially eligible patients in their practices. To be eligible, patients had to communicate in English and have had at least one outpatient visit to HCA in the prior 24 months. Patients were contacted by a letter sent to their home with a description of the study and the opportunity to opt out. Reasons for opting out were collected. We also recruited a convenience sample of BIDMC-affiliated physicians who provide medical care for inpatients. We targeted physicians across a range of medical and surgical specialties as well as hospitalists. Physicians were contacted by email with a description of the study and option to opt out. A priori, we decided to contact patients by phone and physicians by email up to 10 times to see if they were interested in participating in our study. All participants were invited to complete the interview over the phone or in-person in a private office. Recruitment and data collection were conducted simultaneously; invitation letters were mailed in batches, follow-up calls were made in waves, and enrollment and interviews were scheduled as soon as participants agreed and were available.

All participants who enrolled in the study provided verbal consent. A single investigator conducted the semistructured interview (Appendix: https://journals.stfm.org/media/2355/mostofsky_appendix-fmjune19.pdf) for all participants. The Institutional Review Boards at the BIDMC and Harvard T.H. Chan School of Public Health both approved the study protocol. Study data were collected and managed using REDCap electronic data capture tools hosted at Harvard T.H. Chan School of Public Health.[18]

To identify optimal ways of communicating the results of studies based on routinely collected inpatient health information, participants were asked about the results of hypothetical studies (Appendix: https://journals.stfm.org/media/2355/mostofsky_appendix-fmjune19.pdf). The first vignette described a hypothetical study reporting the heightened risk of falling after using sedative hypnotics, and the second vignette described a hypothetical study reporting the lower risk of pneumonia after using oral antiseptics to clean ventilators. After each vignette, participants reviewed several potential descriptions of the results: a description using a risk ratio (short-term incidence rate ratio); a description based on a short-term risk difference; and a description using the number needed to harm or treat. First, participants rated each description in terms of how useful it is for making informed health decisions using a 5-point scale from "not at all" to "extremely." We dichotomized the responses into two categories defined as "not at all," "slightly," or "moderately," vs "very," and "extremely." Second, participants were asked which, if any, of the three descriptions (risk ratio, risk difference, number needed to harm/treat) would be most useful for making informed health decisions. Third, participants were asked which, if any, of the three descriptions (risk ratio, risk difference, number needed to harm/treat) indicates a stronger association between the exposure and outcome.

The third vignette described a hypothetical study reporting that the risk of falls associated with taking a sleeping aid depends on the concomitant use of blood pressure medications, an example of heterogeneity of treatment effects. One interpretation reported multiplicative heterogeneity (variation in the risk ratio associated with sleep aid use among people using or not using blood pressure medications), one reported no multiplicative heterogeneity, one reported additive heterogeneity (variation in the absolute risk difference associated with sleep aid use among people using or not using blood pressure medications), and one reported no additive heterogeneity. Using a 5-point scale from "not at all" to "to a

great extent," patients were asked to rate how likely it is that the finding would impact their decision about accepting or seeking treatment, and physicians were asked how likely it is that the information would impact their treatment plan for such a patient. We dichotomized the responses into two categories defined as "not at all," "not very likely," or "neutral" vs "somewhat" and "to a great extent."

Statistical Analysis

We calculated the frequency of each answer choice for each of the statements to examine which interpretations were considered most informative for patients and physicians. For questions about which of

the results may be most useful for making informed health decisions, we examined whether the responses are similar or not for the two vignettes. For the question regarding which, if any, of the statements indicates a stronger association than any of the other statements, we calculated the frequency of the choices. We used Fisher exact tests to test whether the preferences for the most useful statement and the statement with the strongest association were different between patients and physicians.

Results

The enrollment of study participants is depicted in Figure 1. Among 315

individuals contacted to participate, only 16% (49 patients and 3 physicians) declined to participate. We were unable to contact eight potential participants and there was one incomplete interview. Recruitment and enrollment happened simultaneously. Therefore, once the recruitment goals were met (n=76 patients and 26 physicians), we closed recruitment before we could follow up and enroll the remaining 152 people. Of those who were invited, 102 completed the study (84 phone interviews, 18 interviews in person). Approximately two-thirds of the participants in the study were women, the majority were white and had a high level of education (Table 1). The median

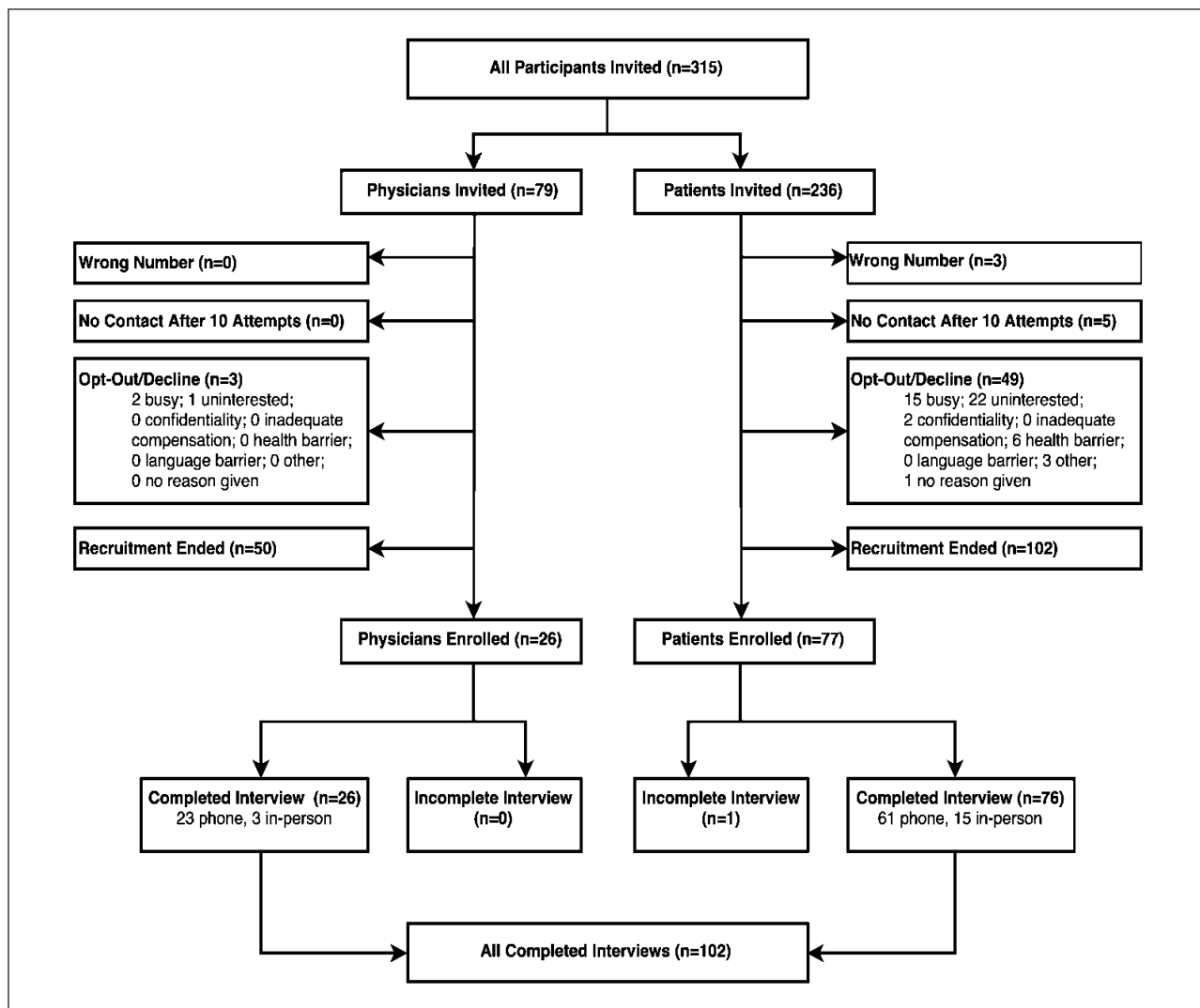**Figure 1: Consort Diagram of Participant Recruitment and Enrollment**

Table 1: Characteristics of Study Participants, Mean±Standard Deviation or n (%)

	Physicians (n=26)	Patients (n=76)
Age (years)	48.03±9.32	54.37±15.07
Women	11 (42.3%)	57 (75.0%)
Ethnicity		
Not Hispanic or Latino	22 (84.6%)	69 (90.8%)
Hispanic or Latino	4 (15.4%)	7 (9.2%)
Race		
American Indian/Alaska Native		1 (1.3%)
Asian	4 (15.4%)	4 (5.3%)
Black or African American	1 (3.8%)	10 (13.2%)
White	21 (80.8%)	57 (75.0%)
More than one race		1 (1.3%)
Unknown/not reported		3 (3.9%)
Highest Level of Education		
Less than high school		1 (1.3%)
High school graduate, diploma or equivalent (GED)		3 (3.9%)
Some college credit, no degree		9 (11.8%)
Trade/technical/vocational training		2 (2.6%)
College graduate		28 (36.8%)
Advanced degree	26 (100.0%)	33 (43.4%)
Occupation		
Medical doctor, hospitalist	12 (46.2%)	
Medical doctor, medical	9 (34.6%)	
Medical doctor, surgical	5 (19.2%)	
Professional/executive/supervisory/technical		37 (48.7%)
Retired		20 (26.3%)
Sales, homemaker, self-employed, clerical		9 (11.8%)
Unemployed		5 (6.6%)
Laborer		3 (4.0%)
Student		1 (1.3%)
Unspecified		1 (1.3%)

time for completing the interview following the consent process was 28 minutes.

Table 2 shows that in response to the vignette about a harmful effect of sedative hypnotics on the immediate risk of falling, physicians and patients did not differ in their rating of the usefulness of the number needed to harm measure (P=.81), whereas patients were more likely than physicians to report that the risk ratio was very or extremely useful (65.8% vs 34.6%, P=.01). Physicians preferred the interpretation of either the number needed to harm or the risk ratio as the most useful measures for making informed health decisions, with 42.4% of physicians selecting these measures as most useful. On the other hand, 59.2% of the patients strongly preferred the risk ratio as the most useful interpretation. In response to the question about which statement indicates the strongest association, patients were most likely to report that the risk ratio provided evidence of the strongest effect of the sedative hypnotics. Though this was also the choice for 46% of the physicians, 39% of them correctly noted that none of the reported interpretations was stronger than any of the others; they simply report different aspects of the association of interest.

In response to the vignette about a beneficial effect of antiseptics to clean ventilators and a lower risk of developing pneumonia, the results were similar to those for the vignette of harmful effects described above. The physicians and patients did not differ in their rating of the usefulness of the number needed to treat measure (P=.14), whereas patients were more likely than physicians to report that the risk difference

Table 2: Preferences for Communicating Results About Reporting Harmful and Protective Effects

	Response	Physicians (n=26)	Patients (n=76)	Fisher Exact P-Value
Vignette 1: Harmful Effects*				
Usefulness of number needed to harm	Very/extremely	8 (30.8%)	27 (35.5%)	.81
Usefulness of risk difference	Very/extremely	3 (11.5%)	24 (31.6%)	.07
Usefulness of risk ratio	Very/extremely	9 (34.6%)	50 (65.8%)	.01
Most useful interpretation	Number needed to harm	11 (42.3%)	16 (21.1%)	.01
	Risk difference	1 (3.9%)	14 (18.4%)	
	Risk ratio	11 (42.3%)	45 (59.2%)	
	None/no response	3 (11.5%)	1 (1.3%)	
Strongest Interpretation	Number needed to harm	3 (11.5%)	8 (10.5%)	.28
	Risk difference	1 (3.9%)	11 (14.5%)	
	Risk ratio	12 (46.2%)	40 (52.6%)	
	None/no response	10 (38.5%)	17 (22.4%)	
Vignette 2: Protective Effects†				
Usefulness of number Needed to yreat	Very/extremely	11 (42.3%)	20 (26.3%)	.14
Usefulness of risk difference	Very/extremely	3 (11.5%)	29 (38.2%)	.01
Usefulness of risk ratio	Very/extremely	13 (50.0%)	66 (86.8%)	<.001
Most useful interpretation	Number needed to treat	11 (42.3%)	3 (4.0%)	
	Risk difference	2 (7.7%)	8 (10.5%)	<.001
	Risk ratio	11 (42.3%)	63 (82.9%)	
	None/no response	2 (7.7%)	2 (2.6%)	
Strongest interpretation	Number needed to treat	3 (11.5%)	5 (6.6%)	
	Risk difference	3 (11.5%)	9 (11.8%)	.01
	Risk ratio	8 (30.8%)	48 (63.2%)	
	None/no response	12 (46.2%)	14 (18.4%)	
Vignette 3: Usefulness of Heterogeneity of Treatment Effects†				
No multiplicative	Somewhat/to a great extent	7 (26.9%)	26 (34.2%)	.63
Multiplicative	Somewhat/to a great extent	22 (84.6%)	68 (89.5%)	.50
No additive	Somewhat/to a great extent	7 (26.9%)	35 (46.1%)	.11
Additive	Somewhat/to a great extent	22 (84.6%)	48 (63.2%)	.74
	No response	0 (0%)	1 (1.3%)	

* The vignette provided potential interpretations for findings from a hypothetical study showing that using sedative hypnotics (drugs that help you sleep) is associated with a heightened risk of falls. See Appendix (https://journals.stfm.org/media/2355/mostofsky_appendix-fmjune19.pdf) for details.

† The vignette provided potential interpretations for findings from a hypothetical study showing that using antiseptics to clean ventilators is associated with a lower risk of developing pneumonia. See Appendix for details.

‡ The vignette provided potential interpretations for findings from a hypothetical study showing that the risk of falls associated with taking a sleeping aid may depend on whether or not it is used with medications for having high blood pressure. See Appendix for details.

(P=.01) and risk ratio (P<.001) measures were very or extremely useful. Physicians and patients differed in their interpretation of the most useful interpretation (P<.001). For example, physicians were equally likely to prefer the interpretation of either the number needed to treat (42.3%) or the risk ratio (42.3%) as the most useful measure for making informed health decisions, whereas patients strongly preferred the risk ratio (82.9%) as the most useful interpretation. Once again, patients reported that the risk ratio provided evidence of the strongest protective effects. However, though this was also the choice for 31% of the physicians, 46% of them correctly noted that none of the reported

interpretations was stronger than any of the others (Table 2).

In the vignette about potential heterogeneity of treatment effects, we described findings reporting that the risk of falls associated with taking a sleeping aid may depend on concomitant use of antihypertensive medications. Participants were asked to report how likely their decision about taking or prescribing a sleeping aid would depend on whether or not the individual is also taking high blood pressure medication. In the first interpretation, there was no evidence of multiplicative heterogeneity of treatment effects; in the second interpretation, there was multiplicative heterogeneity of treatment effects; in the third interpretation, there was no evidence of additive heterogeneity of treatment effects; in the fourth interpretation, there was additive heterogeneity of treatment effects. Physicians were equally likely to report that their decision would be impacted by multiplicative or additive heterogeneity of treatment effects (84.6% for both), whereas patients were more likely to report that their decision would be influenced by multiplicative heterogeneity of treatment effects (89.5%; Table 2).

Discussion

With the rapid increase in the availability of information collected in electronic health records, there is an enormous opportunity to conduct studies addressing the potential risks and benefits of health characteristics, medications, and procedures. However, the information is only useful if the results can be communicated in a way that helps patients and physicians make informed decisions.

In this study, we aimed understand the best way to communicate research findings. Despite the importance of considering absolute risks and benefits of treatment options, the physicians were equally likely to prefer the interpretation of either the number needed to treat or harm, or the risk ratio as the most useful

measure for making informed health decisions, and the patients strongly preferred the risk ratio as the most useful interpretation. Concordant with previous studies,[2,13,19] participants perceived changes in risk as larger when they were presented using multiplicative measures such as the risk ratio than when presented as additive measures such as the risk difference or the number needed to treat or harm.

There are some limitations to our study that warrant discussion. This moderately sized study is based on a convenience sample comprised of interested patients and physicians at BIDMC. Therefore, the numeracy of our participants may not reflect those of other more diverse populations. In our questions about the different measures, we did not elicit responses about how the participants processed and understood the different interpretations of the results. It is possible that preferences were also affected by the specific exposures and outcomes in our vignette examples. The health care experience for patients is also likely to vary by the sampling frame for the study. Involving patients in shared medical decision-making may need to be tailored to their educational background, health literacy,[4] and age,[20] and the choice may depend on issues surrounding inpatient vs outpatient care. Our sample of stakeholders includes people at risk of being hospitalized at BIDMC, but not necessarily people who have been hospitalized. Our participant stakeholders may have different preferences than those who are currently facing choices required of inpatient care. They may differ from patients who receive care in suburban practices, concierge practices, or community health centers. We did not examine preferences regarding presenting the risk in each group separately for exposed and unexposed rather than their additive or multiplicative associations. We also did not examine the trade-offs between risks and benefits of treatment options or the time frame of potential

risk that should be used for making informed decisions. However, based on these detailed interviews with 102 participants, we were able to gain insight on their preferences for communicating study findings.

The results of our study highlight one of the complexities of the decision-making process. Although additive measures provide useful information about health risks and benefits, their utility is limited by the confusion over the difference between additive and multiplicative measures and how this impacts conclusions about heterogeneity of treatment effects. Understanding the cognitive process of interpreting health findings may help guide methods for effective health communication.[12, 21-23] Given the importance of involving patients and other stakeholders in making informed health decisions, clear and relevant descriptions of study findings are critical.

CORRESPONDING AUTHOR: Address correspondence to Dr Elizabeth Mostofsky, Harvard T.H. Chan School of Public Health, Department of Epidemiology, Kresge Building, Room 505-B, 677 Huntington Ave, Boston MA 02115. 617-432-4023. emostofs@mail.harvard.edu.

References

1. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9.

2. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ. 2003;327(7417):741-744.

3. Paling J. Strategies to help patients understand risks. BMJ. 2003;327(7417):745-748.

4. Institute of Medicine. 2014. Health Literacy and Numeracy: Workshop Summary. Washington, DC: The National Academies Press. .

5. Girerd N, Rabilloud M, Pibarot P, Mathieu P, Roy P. Quantification of Treatment Effect Modification on Both an Additive and Multiplicative Scale. PLoS One. 2016;11(4):e0153010.

6. Ferreira JP, Coiro S, Girerd N. A Systematic Assessment of Absolute Treatment Effect. Am J Cardiol. 2015;116(5):829-831.

7. Diamond GA, Kaul S. On reporting of effect size in randomized clinical trials. Am J Cardiol. 2013;111(4):613-617.

8. Barratt A, Wyer PC, Hatala R, et al. Tips for learners of evidence-based medicine: 1: Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ. 2004;171(4):353-8.

9. Bolland MJ, Grey A. Data openness. Reporting of absolute risk. BMJ. 2010;341(nov16 4):c6331.

10. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. BMJ. 2010;341(oct12 2):c4830.

11. Norman PE. Data openness. Absolute v relative risk reduction. BMJ. 2010;341(nov16 4):c6333.

12. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull. 2009;135(6):943-973.

13. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. J Natl Cancer Inst. 2011;103(19):1436-1443.

14. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med. 1992;117(11):916-921.

15. Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. Cochrane Database Syst Rev. 2011;(3):CD006776.

16. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: absolute risk reduction, relative risk reduction, and number needed to treat. Perspect Clin Res. 2016;7(1):51-53.

17. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012;41(2):514-520.

18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-381.

19. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. Psychol Sci Public Interest. 2007;8(2):53-96.

20. Tsai TI, Yu WR, Lee SD. Is health literacy associated with greater medical care trust? Int J Qual Health Care. 2018.

21. Graffeo M, Polonio L, Bonini N. Individual differences in competent consumer choice: the role of cognitive reflection and numeracy skills. Front Psychol. 2015;6:844.

22. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. Ann Behav Med. 2008;35(3):261-74.

23. Garcia-Retamero R, Okan Y, Cokely ET. Using visual aids to improve communication of risks about health: a review. ScientificWorldJournal. 2012;2012:562637.