practice performance? And what are the implications for both measurement and performance evaluation?

Firstly, the variance in satisfaction scores is not surprising given the multidimensional nature of health care and patient satisfaction. Although satisfaction is seen as a judgment about whether expectations were met, it is influenced by varying standards, different expectations, the patient's disposition, time since care, and previous experience.[2] None the less, qualitative research shows that patients will give positive satisfaction ratings even in the face of a negative experience unless they believe that the poor care is under the direct control of the person they are evaluating.[3 4] For example, they may be unhappy about hurried communication with their doctor but still give an adequate rating because they attribute this to time constraints not a lack of intrinsic skills. Consequently, positive satisfaction ratings include both true positives and false positives. This compromises sensitivity in a diagnostic test and by the same token reduces the precision of satisfaction ratings. In contrast, negative satisfaction ratings tend to be truly negative (or highly specific in the analogy of diagnostic accuracy) and reflect important incidents, such as a lack of respect or medical errors.[4 5] The implication is that the representation of satisfaction and satisfaction ratings needs to be changed. It is better to report the proportion of patients who are less than totally satisfied rather than the average satisfaction. High satisfaction ratings indicate that care is adequate not that it is of superior quality; low ratings indicate problems and should not be masked by reporting average scores.

Secondly, a defining characteristic of primary care is its high degree of variety and variance, even within the practice of one doctor.[6 7] On a technical note, it is important to remember that analytical modelling that separates the variance into practice, doctor, and patient levels cannot separate variance between patients from random error. Part of this random error comes from the variation within practices and within doctors, which is to be expected, given the complexity of primary care. It is not surprising that such complexity can be only partially captured by a short questionnaire about experience and satisfaction. Despite this, patient assessments of health care work surprising well. Salisbury and colleagues show that assessments of access explain more variance between practices than they do between doctors, which makes sense for an attribute related to organisational arrangements. Conversely, assessments of communication explain more variance between doctors than between practices. Other studies have also found that patient assessments appropriately detect more variance between practices for organisational attributes and between doctors for personal care attributes.[8 9] The implication is that the differences between practices and between doctors seen in the current analytical models underestimate the true differences that occur at the practice and doctor levels, and although Salisbury and colleagues are right in advocating prudence in interpreting small differences between practices, we can be confident that statistically significant differences are real and clinically relevant.

Thirdly, these results have implications for improving the science of measurement. Although it is difficult to measure patients' perceptions of health care, it is most appropriate that patients should assess the interpersonal dimension of quality of care because they are the ones to whom we are ultimately accountable. It is therefore crucial that patient surveys are refined to maximise precision and minimise bias. The research community needs to develop and refine robust and comparable measures, bearing in mind that deficiencies in the measurement of satisfaction are more common in newly devised instruments.[4]

Measures of patient satisfaction need to be refined, but they are not hopelessly flawed. When they detect problems, these are real and important. They should be presented in a way that highlights the informative negative assessments, and they need to be combined with reports (such as experience) of components that can be benchmarked to recognised best practices.

1 Salisbury C, Wallace M, Montgomery A. Patient experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ* 2010;341:c5004.
2 Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess* 2002;6:1-244.
3 Schneider H, Palmer N. Getting to the truth? Researching user views of primary health care. *Health Policy Plan* 2002;17:32-41.
4 Collins K, O'Cathain A. The continuum of patient satisfaction—from satisfied to very satisfied. *Soc Sci Med* 2003;57:2465-70.
5 Taylor B, Marcantonio ER, Pagovich O, Carbo A, Bergmann M, Davis RB, et al. Do medical inpatients who report poor service quality experience more adverse events and medical errors? *Med Care* 2008;46:224-8.
6 Love T, Burton C. General practice as a complex system: a novel analysis of consultation data. *Fam Pract* 2005;22:347-52.
7 Katerndahl DA, Wood R, Jaén CR. A method for estimating relative complexity of ambulatory care. *Ann Fam Med* 2010;8:341-7.
8 Haggerty JL, Pineault R, Beaulieu M-D, Brunelle Y, Gauthier J, Goulet F, et al. Practice features associated with patient-reported accessibility, continuity and coordination of primary health care. *Ann Fam Med* 2008;6:116-23.
9 Rodriguez HP, Scoggins JF, von Glahn T, Zaslavsky AM, Safran DG. Attributing sources of variation in patients' experiences of ambulatory care. *Med Care* 2009;47:835-41.

Misleading communication of risk

Editors should enforce transparent reporting in abstracts

Cite this as: *BMJ* 2010;341:c4830
doi: 10.1136/bmj.c4830

In 1996 a review of mammography screening reported in its abstract a 24% reduction of breast cancer mortality[1]; a review in 2002 claimed a 21% reduction.[2] Accordingly, health pamphlets, websites, and invitations broadcast a 20% (or 25%) benefit.[3] Did the public know that this impressive number corresponds to a reduction from about five to four in every 1000 women, that is, 0.1%? The answer is, no. In a representative quota sample in nine European countries, 92% of about 5000 women overestimated the benefit 10-fold, 100-fold, and more, or they did not know.[4] For example, 27% of women in the United Kingdom believed that out of every 1000 women who were screened, 200 fewer would die of breast cancer. But it is not only patients who are misled. When asked what the

Gerd Gigerenzer director
sekgigerenzer@mpib-berlin.mpg.de
Odette Wegwarth research scientist
Markus Feufel postdoctoral fellow, Harding Center for Risk Literacy, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Competing interests: None declared.
Provenance and peer review: Commissioned; not externally peer reviewed.

"25% mortality reduction from breast cancer" means, 31% of 150 gynaecologists answered that for every 1000 women who were screened, 25 or 250 fewer would die.[3]

In 1995, the UK Committee on Safety of Medicines issued a warning that third generation oral contraceptive pills increased the risk of potentially life threatening thrombosis twofold. The news provoked great anxiety, and many women stopped taking the pill, which led to unwanted pregnancies and abortions––some 13 000 additional abortions in the next year in England and Wales—and an extra £46m (€55m; $71m) in costs for the NHS.[5] Yet how big was the twofold risk? The studies revealed that for every 7000 women who took the earlier, second generation pills, one had a thrombosis, and this number increased to two in women who took third generation pills. The problem of misleading reporting has not gone away. In 2009, the *BMJ* published two articles on oral contraceptives and thrombosis; one made the absolute numbers transparent in the abstract,[6] whereas the other reported that "oral contraceptives increased the risk of venous thrombosis fivefold."[7]


DAVID PAGE PHOTOGRAPHY/ALAMY

These two examples illustrate a general point. Absolute risks (reductions and increases), such as from one to two in 7000, are transparent, while relative risks such as "twofold" provide incomplete and misleading risk information.[3 8] Relative risks do not inform about the baseline risk—for example, whether twofold means from one to two or from 50 to 100 in 7000—and without this information, people overestimate benefits or harms.[3 9] In the case of the pill scare, the losers were women, particularly adolescent girls, taxpayers, and the drug industry. Reporting relative risks without baseline risk is practised not only by journalists because big numbers make better headlines or by health organisations because they increase screening participation rates. The source seems to be medical journals, from which figures spread to press releases, health pamphlets, and the media.

An analysis of the articles published in the *Annals of Internal Medicine*, *BMJ*, *JAMA*, *Journal of the National Cancer Institute*, *Lancet*, and the *New England Journal of Medicine*, 2003-4, showed that 68% (150/222) failed to report the underlying absolute risks in the abstract. Among those, about half did report the absolute risks elsewhere in the article, but the other half did not.[10] Similarly, an analysis of 119 systematic reviews in *BMJ*, *JAMA*, and *Lancet* from 2004 to 2006 showed that every second article discussed only relative risks or odds ratios.[11]

Conveying relative risks without baseline risk is the first "sin" against transparent reporting. The second is mismatched framing—reporting benefits, such as relative risk reductions, in big numbers and harms, such as absolute risk increases, in small numbers.[3] If we use the example of a treatment that reduces the probability of getting disease A from 10 to five in 1000, whereas it increases the risk of disease B from five to 10 in 1000, authors who use mismatched framing would report the benefit as a 50% risk reduction and the harm as an increase of five in 1000; that is, 0.5%. Medical journals permit mismatched framing. One in three articles in the *BMJ*, *JAMA*, and *Lancet* from 2004 to 2006 used mismatched framing when both benefits and harms were reported.[11]

Have editors since stopped non-transparent reporting? To check the current situation, we examined the abstracts of all free accessible research articles published in the *BMJ* in 2009 that reported drug interventions. Of the 37 articles identified, 16 failed to report the underlying absolute numbers for the reported relative risk measures in the abstract. Among these, 14 reported the absolute risks elsewhere in the article, but two did not report them anywhere. Moreover, absolute risks or the number needed to treat (NNT) were more often reported for harms (10/16) than for benefits (14/27).

How can those who are responsible for accurate communication of risk do better? And who should be monitoring them to ensure that they do? Steps can be taken to improve the transparency of risk communication.[12] Firstly, editors should enforce transparent reporting in journal abstracts: no mismatched framing, no relative risks without baseline risks, and always give absolute numbers such as absolute risks and NNT.

Secondly, institutions that subscribe to medical journals could give journal publishers two years to implement the first measure and, if publishers do not comply, cancel their subscriptions.

Thirdly, writers of guidelines, such as the CONSORT statement, should stipulate transparent reporting of benefit and harms in abstracts.

1 Larsson LG, Nyström L, Wall S, Rutqvist L, Andersson I, Bjurstam N, et al. The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. *J Med Screen* 1996;3:129-32.
2 Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-19.
3 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients to make sense of health statistics. *Psychol Sci Public Interest* 2007;8:53-96.
4 Gigerenzer G, Mata J, Frank R. Public knowledge of benefits of breast and prostate cancer screening in Europe. *J Natl Cancer Inst* 2009;101:1216-20.
5 Furedi A. The public health implications of the 1995 "pill scare." *Hum Reprod Update* 1999;5:621-6.
6 Lidegaard Ø, Løkkegaard E, Svendsen AL, Agger C. Hormonal contraception and risk of venous thromboembolism: national follow-up study. *BMJ* 2009;339:b2890.
7 Van Hylckama Vlieg A, Helmerhorst FM, Vandenbroucke JP, Doggen CJM, Rosendaal FR. The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* 2009;339:b2921.
8 Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;327:741-4.
9 Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making* 2007;27:638-54.
10 Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ* 2006;333:1248-52.
11 Sedrakyan A, Shih C. Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. *Med Care* 2007;45:523-8.
12 Gigerenzer G, Gray JAM. Launching the century of the patient. In: Gigerenzer G, Gray JAM, eds. *Better doctors, better patients, better decisions: envisioning healthcare 2020*. MIT Press [forthcoming].