Improving Depiction of Benefits and Harms

Analyses of Studies of Well-Known Therapeutics and Review of High-Impact Medical Journals

Artyom Sedrakyan, MD, PhD,*†‡ and Chuck Shih, MHS*

Abstract: The issues of weighing benefits and harms and of shared decision-making have become increasingly important in recent years. There is limited knowledge and lack of adequate data on the most transparent method of communicating the information. In this article we discuss examples of communicating benefits and harms for well-known therapeutics, illustrating that relative risk estimates are not helpful for communicating the chance of experiencing adverse events. In addition, we show that asymmetric presentation of the data for benefits and harms is likely to bias toward showing greater benefits and diminishing the importance of the harms (or vice versa). We also present preliminary results of a brief review of high-impact medical journals that show limitations of current systematic reviews. In the review we found that every second published study does not discuss frequency data and 1 in 3 studies that report information on both benefits and harms does not report information in the same metric. We conclude that consistently depicting benefit and harm information in frequencies can substantially improve the communication of benefits and harms. Investigators should be requested to provide frequency data along with relative risk information in the publication of their scientific findings. Currently, even in the highest impact medical journals, evidence of benefits and harms is not consistently presented in ways that facilitate accurate interpretation.

Key Words: risk communication, benefits and harms, patient education, informed choice, systematic review

(*Med Care* 2007;45: S23–S28)

Solving a problem simply means representing it so as to make the solution transparent.
Herbert A. Simon
The Sciences of the Artificial

The issues of weighing the benefits and harms of therapeutics and shared decision-making has become increasingly important in recent years. A recent Institute of Medicine (IOM) forum on understanding the benefits and harms of pharmaceuticals[1] highlighted the lack of adequate data and generally limited knowledge about what would be the most transparent method of communicating medical information to people. This is hardly a surprise, as the science of evidence communication is relatively new. Its importance grew, at least partially, as a result of the rapid growth of medical information in the last 2 decades. Two more recent developments were critical for the renewed interest in advancing this science: first, the need for shared medical decision-making; second, appreciation of the principles of evidence-based medicine.

The goal of evidence communication is informed decision-making. Hence, it is important that people receive information in an understandable format and be able reflect on the trade-offs involved with the therapy.[2] Unless the information is presented as it applies to the individual patient, a choice for one or another type of therapy cannot be made. In essence, medical-decision making has a mathematical core and both researchers and physicians have a mission to communicate that core in such a way that it is best understood by a patient. In this context, the issue of weighing marginal benefit and marginal harm depends on and is potentially secondary to the first step: how to represent numeric information on benefits and harms. Traditionally, the most common ways of communicating information on benefits and harms were qualitative terms such as "high or low risk" or relative terms such as "percent reduction" or "percent increase" in the occurrence of the events. In the late 80s and early 90s, frequency measures such as numbers needed to treat (NNT) were introduced and hailed as a useful way to communicate scientific information.[3,4] Although these terms were generally accepted by the medical community, there has recently been controversy about physicians' and patients' ability to respond to information presented in this format. The opponents of NNT report that physicians and patients have trouble understanding these terms[5,6] whereas advocates attribute these reports to "unusual framing" when communicat-

ing information in frequency terms.[7] The debate highlights the uncertainty about the most comprehensible way of communicating the information to physicians and patients. Gigerenzer and colleagues[8,9] provide some compelling arguments for using absolute or marginal frequency information (events per x treated or additional events per x treated) and have shown that relative terms are likely to be severely misunderstood by both patients and clinicians. However, the attractive statistical features of relative terms are still used to justify their use, and the assumption is that physicians will need to make further efforts to translate relative terms into understandable format. This is likely to be a naive assumption, given the limited time and inadequate training of physicians in risk communication. Furthermore, the inability of both providers and patients to understand numeric information is a "collective innumeracy"[7] and is an impediment to informed decision making.

In this article, we sought to concentrate on issues of evidence communication and provide illustrative examples to support the use of the frequency information. First, we discuss examples of well-known therapeutics and their association with stroke or cerebrovascular events in both frequency and relative terms. Furthermore, we discuss examples of extremely confusing asymmetric presentation of benefits and harms. Finally, we report the preliminary findings of a brief review of 3 leading medical journals [*Journal of American Medical Association (JAMA), British Medical Journal (BMJ), and Lancet*] to evaluate the reporting of benefits and harms in systematic reviews.

EXAMPLES OF WELL-KNOWN THERAPEUTICS

Relative Risk (Risk Ratio, Odds Ratio) Versus Frequencies (Chance)

Relative risk estimates are meaningless for communicating the chance of experiencing an event (either benefit or harm).

Oral Contraceptives and Stroke

The risk communication tools recommended by the Association of Reproductive Health Professionals rely heavily on frequency data to educate women about the benefits and harms of oral contraceptives (OC).[10] For example, to communicate cardiovascular harm, educational materials refer to studies that report frequency terms such as "4 versus 2 women will experience heart attack out of 1 million

OC users (for women <35 years old)" rather than using relative terms.[11]

To evaluate stroke harms, a landmark systematic review reported the harm information as 4.1 additional strokes per 100,000 women taking OC or 1 additional ischemic stroke per year per 24,000 women using OC.[12] The study also reported a relative risk estimate of 1.93, which can be interpreted as a 93% increased risk of stroke (or an increase by a factor of 1.93).

The 93% increase might sound frightening to many women, whereas the chance of 1 in 24,000 is hardly considered troubling (Table 1). Internet searches on the risk of experiencing a harmful event associated with OC usually yield similar information presented in frequencies. As an example, the first hit in the Google search for "risk of oral contraceptives" will yield a website explaining the chances of stroke or other harms in frequency terms rather than relative terms.[13] In this instance, the public health impact of the additional chance of stroke needs to be weighed against the benefits.

Phenylpropanolamine and Stroke

Phenylpropanolamine (PPA), an appetite suppressant used mostly by women, was considered a "lifestyle drug" to lose weight. It was also a component of many cough-cold remedies. As early as 1984, information was emerging about possible hemorrhagic stroke harm associated with PPA.[14]

A landmark study funded by the pharmaceutical industry was designed and conducted by a group of independent researchers from Yale University with help from the US Food and Drug Administration (FDA). The study was published in the *New England Journal of Medicine (NEJM)*. It was found that any PPA use was associated with 2 times higher odds of stroke; use as appetite suppressants was associated with 16.6 times higher odds of stroke compared with no use of this medication.[15] This case–control study involved large numbers of stroke and control patients. Because this was a case–control study, we note that analysts could not use their data to estimate the chance of experiencing stroke. The chance of experiencing stroke was almost never discussed with this drug and only odds ratios were communicated. The FDA public health advisory said that PPA was a likely cause of 200–500 hemorrhagic strokes annually and said that chances of a stroke were "low."[16] Although the exact estimates were not given, millions of women (as high as 10 million) took PPA annually in the years before publication of this study. If 200–500 strokes were to be observed for 10 million users annually, the additional annual chance of

---

TABLE 1. Relative Risk, Number Needed to Harm, Frequency Information, and Public Health Impact for Selected Drugs and Their Association With Stroke

Therapy	Relative Risk	No. Needed to Harm (NNH)	Additional Annual Chance of Stroke per 100,000 Persons	Public Health Impact Annually
Oral contraceptives	1.93	24,000 (study)	4 in 100,000	440 strokes
Phenylpropanolamine	16.6	20,000–500,000 (calculated)	<1 to 5 in 100,000	<50 to 500 strokes
Hormone replacement therapy	1.29	160 (study)	>92 in 100,000	>13,675 strokes
Atypical antipsychotics	7.8*	28 (study)	3750 in 100,000	7000 strokes

*Estimated from 0 (167) vs. 6 (170) events using Peto method.

experiencing a stroke can be calculated as 1 in 20,000 to 50,000 women. One of the authors of the Yale report, Dr. Lawrence Brass, put his estimate for the chance of hemorrhagic stroke as "perhaps 1 in 500,000" users.[17]

When compared with oral contraceptives, PPA is likely to be 10 times less harmful in terms of stroke in the least conservative scenario and probably equally harmful in the most conservative scenario. However, if one looks at the odds ratio information, then a very different conclusion is reached (Table 1). Interestingly, the risk information (over 16 times higher odds) presented in the *NEJM* article left a strong impression on the public and many researchers. In this instance, the drug was taken off the market. Availability of safer alternatives was at least partially related to this decision. Of note is the fact that alternatives were not as commonly used at the time and their safety was even more difficult to establish.

Hormone Replacement Therapy and Stroke

Hormone replacement therapy (HRT) was quite popular in past decades, although its use is reportedly decreasing since the publication of the Women's Health Initiative (WHI).[18,19] However, the potential to reduce menopausal symptoms and the medicalization of menopause make HRT appealing to many women. The initial epidemiological studies reported additional benefits of HRT, including cardiovascular benefits, which were widely debated. Randomized controlled trials (RCTs) did not confirm any cardiovascular benefits but instead reported an increased risk of breast cancer. In addition, the latest systematic review has shown that HRT is associated with a 29% higher risk of stroke (a factor of 1.29) compared with no treatment.[20] Again, when looking at the relative risk information alone, the increased chance of harm might be considered low when compared with oral contraceptives (93% increase in stroke) or PPA (16.6-fold increase). However, when estimating the frequency data, one finds that HRT is likely to be over 20 times more harmful than either OC or PPA. The additional annual chances of experiencing a stroke for a middle-aged woman taking HRT are over 92 per 100,000 (Table 1). This phenomenon is related to higher overall occurrence of stroke in postmenopausal women. Accordingly, even a small relative risk increase in stroke translates into a much higher chance of stroke in this group of women.

The public health impact of HRT is hard to ignore based on these harm data, and serious questions need to be asked about the net health benefits (the balance of benefits and harms). Again, communication of frequency information rather than relative risks is critical to debate. The chance of experiencing harms and benefits is not consistently communicated to women. Proper communication might be the key in decision-making about the use or marketing of this therapy in future.

Emerging Evidence on Atypical Antpsychotics

A recent systematic review reported the benefits and harms of atypical antipsychotics.[21] The review summarized 5 RCTs that discussed harms. Only 1 study reported information on stroke, finding 7 strokes in 170 patients assigned to antipsychotics and zero in 167 assigned to placebo.[22] Neither the original study nor the systematic review calculated relative risk or statistical significance. Only numeric information on this harm was presented. Using the Peto method to calculate the odds ratio yields an estimate for harm of 7.8 ($P < 0.01$). As all of the strokes in this study occurred within the first months of taking the medication, the annual occurrence of stroke would be at least comparable to this estimate and might be even higher. The chance of stroke in the eligible population can be as high as 1 in 28 people (Table 1). Although the estimates might change when more information from other studies becomes available, the public health impact is too large to ignore, and the frequency data will trigger more thorough studies of the benefits and harms.

Asymmetric Communication of Benefits and Harms

Benefits may be presented in relative terms but harms presented in terms of frequencies (or vice versa). This is very likely to asymmetrically bias toward showing greater benefits and diminishing the importance of the harms (or vice versa).

Recombinant Tissue Plasminogen Activator Administration After Stroke

Intravenous recombinant tissue plasminogen activator (rt-PA) is considered useful for the treatment of ischemic strokes. A pooled analysis of 6 RCTs published in 2004 discussed early and late rt-PA administration.[23] The study reported that the "odds of the favorable outcome" increase the sooner the rt-PA is administered (odds ratios of 2.8, 1.6, 1.4, and 1.2 for the administration of rt-PA at <90, 91–180, 181–270, and 271–360 minutes after stroke, respectively). However, harms are presented in frequencies; for example as "82 (5.9%) rt-PA patients and 15 (1.1%) in controls" experienced serious hemorrhage. One can think of this as an odds ratio of >5.0 for the sake of consistency, but this is hardly helpful anyway. One cannot learn about the balance of the benefits and harms without having the frequency information for both benefits and harms. The conclusion states, "The sooner that rt-PA is given to stroke patients, the greater the benefit, especially if started within 90 minutes. Our results suggest a potential benefit beyond 3 hours, but this potential might come with some risks." There is not enough information presented in the article for a reader to reach a similar conclusion. From the given information, one can discuss the harms as 48 additional serious hemorrhages per 1000 patients administered rt-PA. Similarly, number needed to harm can be calculated as 21 (100%/(5.9–1.1%), which means that 1 additional serious hemorrhage occurs per 21 treated patients. However, a similar calculation for benefits is not possible, and this lack of transparency might bias an informed reader toward thinking that benefits do not outweigh harms.

Aggrenox (Combination of Dipyridamole and Aspirin) for Secondary Stroke Prevention

Evidence on the stroke reduction associated with Aggrenox administration is frequently linked to the European Stroke Prevention Study 2, published in 1996.[24] The FDA based its approval decision on the same study,[25] stating that, "Aggrenox reduced the risk of stroke by 36.8% and the

cumulative risk of stroke and death by 24.2% compared with placebo." The article highlights relative risk reduction in stroke in the abstract and in the main text. However, the abstract reports no numeric data for serious adverse events and only a qualitative harm statement for all-site bleeding and gastrointestinal bleeding is reported with significantly more events in patients who received Aggrenox when compared with aspirin or dipyridamole.

Aggrenox data is not consistently presented in the article to judge about benefits and harms of Aggrenox when compared with aspirin alone or dipyridamole alone. A careful examination of the article might show that Aggrenox, compared with aspirin alone, is associated with 30 additional stroke reductions per 1000 patients, but also 3 additional increased deaths per 1000 patients treated over 2 years. The frequency data also shows approximately 4 additional severe or fatal bleeding episodes with Aggrenox than with aspirin alone. If these harms had been clarified and transparently presented in the article, the research community could have had a different opinion about the net health benefit of this drug.

EVALUATING BENEFITS AND HARMS DEPICTION IN SYSTEMATIC REVIEWS PUBLISHED IN 3 HIGH-IMPACT JOURNALS

We sought to determine whether any of the following problems occurred in 3 high-impact journals that publish systematic reviews:

1. Frequency data not reported;
2. Benefits and harms not presented in the same metric (ie, relative risks for benefits and frequencies for harms).

We hand searched *BMJ, JAMA*, and *Lancet* from January 2004 through May 2006 for all systematic reviews of therapies. We found 216 studies; of these, we excluded 52 prevalence studies, 21 studies of diagnostic tests, and 24 studies of nontherapeutic interventions (ie, diet). We included 119 studies of therapeutic interventions; of these, 64% were drug therapy studies, 19% studies of noninvasive interventions, 10% studies of other therapeutic invasive interventions, and only 7% studies involving surgical treatments.

We found that both benefits and harms were reported in 55% of studies, 34% of the studies reported only benefits, and 11% reported only harms.

Figure 1 summarizes the most important findings of our preliminary review. Roughly every second study published in high-impact medical journals does not discuss frequency data. For the most part, the articles discuss relative estimates (relative risks or odds ratios; Fig. 1). Further, 1 in 3 studies reporting information on both benefits and harms does not present the benefit and harm information in the same metric. In most cases, relative risk is reported for benefits and frequency data are reported for harms.
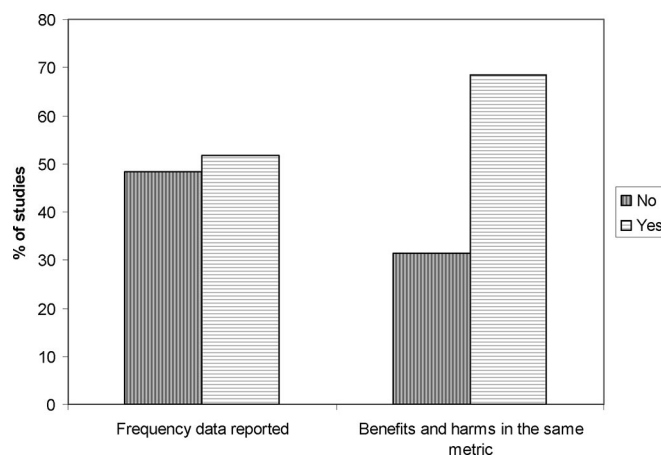

FIGURE 1. Numeric communication in studies published in high-impact medical journals.

CONCLUSIONS

Consistently depicting benefit and harm information in frequencies in technology assessments, evidence reviews, and individual studies can substantially improve benefit and harm communication. Investigators should be asked to provide frequency data along with relative risk information when publishing their scientific findings.

Currently, evidence of benefits and harms is not consistently presented in ways that facilitate accurate interpretation even in the highest impact medical journals.

ACKNOWLEDGMENTS

REFERENCES

1. IOM. Forum on Drug Discovery, Development, and Translation Workshop 2: Understanding the Benefits and Risks of Pharmaceuticals. 2006. Available at: http://www.iom.edu/CMS/3740/24155/32576.aspx. Accessed November 22, 2006.
2. McNutt RA. Shared medical decision making: problems, process, progress. *JAMA*. 2004;292:2516–2518.
3. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318: 1728–1733.
4. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995;310:452–454.
5. Halvorsen PA, Kristiansen IS. Decisions on drug therapies by numbers needed to treat: a randomized trial. *Arch Intern Med*. 2005;165:1140–1146.
6. Sheridan SL, Pignone MP, Lewis CL. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *J Gen Intern Med*. 2003;18:884–892.
7. Elmore JG, Gigerenzer G. Benign breast disease—the risks of communicating risk [comments]. *N Engl J Med*. 2005;353:1856–1858.
8. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ*. 2003;327:741–744.
9. Gigerenzer G. *Reckoning With Risk*. 1st ed. London, England: Penguin Books; 2002.

10. ARHP. A.O.R.H.P. Tools for Communicating Risk About Hormonal Contraceptives. 2007. Available at: http://www.arhp.org/healthcareproviders/cme/onlinecme/RiskProjectCP/tools.cfm. Accessed March 16, 2007.

11. Farley TM, Collins J, Schlesselman JJ. Hormonal contraception and risk of cardiovascular disease. An international perspective. *Contraception*. 1998;57:211–230.

12. Gillum LA, Mamidipudi SK, Johnston SC. Ischemic stroke risk with oral contraceptives: a meta-analysis. *JAMA*. 2000;284:72–74.

13. FHI. F.H.I. What Are the Benefits and Risks of Combined Oral Contraceptives? Reproductive Health Fact Sheet 2006. Available at: http://www.fhi.org/en/RH/Pubs/factsheets/ocriskben.htm. Accessed October 27, 2006.

14. Jick H, Aselton P, Hunter JR. Phenylpropanolamine and cerebral haemorrhage. *Lancet*. 1984;1:1017.

15. Kernan WN, Viscoli CM, Brass LM, et al. Phenylpropanolamine and the risk of hemorrhagic stroke. *N Engl J Med*. 2000;343:1826–1832.

16. Phenylpropanolamine. 2006. Available at: http://en.wikipedia.org/wiki/Phenylpropanolamine#_note-2. Accessed October 27, 2006.

17. Dabela A. Yale Study Parks FDA Review of Cold Drugs. The Yale Herald News Online 2000. Available from http://www.yaleherald.com/archive/xxx/2000.10.27/news/index.html. Accessed October 27, 2006.

18. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.

19. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *JAMA*. 2002;288:321–333.

20. Bath PM, Gray LJ. Association between hormone replacement therapy and subsequent stroke: a meta-analysis [review]. *BMJ*. 2005.330:342.

21. Lee PE, Gill SS, Freedman M, et al. Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia: systematic review. *BMJ*. 2004;329:75.

22. Brodaty H, Ames D, Snowdon J, et al. A randomized placebo-controlled trial of risperidone for the treatment of aggression, agitation, and psychosis of dementia. *J Clin Psychiatry*. 2003;64:134–143.

23. Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet*. 2004;363:768–774.

24. Diener HC, Cunha L, Forbes C, et al. European stroke prevention study, part 2: dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *J Neurol Sci*. 1996;143(1–2):1–13.

25. FDA. FDA Approves New Drug to Reduce Risk of Stroke. 1999. Available at: http://www.fda.gov/bbs/topics/ANSWERS/ANs00985.html. Accessed October 27, 2006.

26. Greendale GA, Reboussin BA, Hogan P, et al. Symptom relief and side effects of postmenopausal hormones: results from the postmenopausal estrogen/progestin interventions trial. *Obstet Gynecol*. 1998;92:982–988.

27. Nikolajevic-Sarunac J, Henry DA, O'Connell DL, et al. Effects of information framing on the intentions of family physicians to prescribe long-term hormone replacement therapy. *J Gen Intern Med*. 1999;14:591–598.

28. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. *Lancet* 1997;350:1047–1059.

29. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med*. 1992;117:916–921.

APPENDIX A

An Example and Possible Solution to Numeric Problems in a Hypothetical Instance of Doctor–Patient Interaction

Consider the following example of doctor–patient interaction after the evidence for HRT changed substantially.[18,19]

Clara S, a widowed 48-year-old mother of 2, had been using HRT for 2 months when she learned about possible harms associated with HRT. Clara was quite concerned about this and made an appointment with her clinician the next day.

Dr. X sounded quite knowledgeable. She explained that a lot of information about HRT came from large observations of women like Clara and reminded her about the benefits:

1. Women taking HRT are likely to have about a 50% reduction in menopausal symptoms such as hot flashes,[26]
2. HRT is associated with a decreased occurrence of hip fracture that can range from 15% to 43%.[27]

Then Dr. X shared a finding of a possible risk; 20–23 women out of 1000 taking HRT will develop breast cancer after 10 years of HRT use, as opposed to about 18 out of 1000 not taking HRT.[28] In addition, HRT is not associated with cardiovascular benefits, as was previously thought, and in fact might be associated with a small increased risk of these events.[19] Clara was not convinced that her hot flashes and sweating were down by 50% and thought they might get better later on if she continued on HRT. Therefore, the 50% reduction was very good news. What was on the other end? The small risk of breast cancer and possible cardiovascular problems? She was not happy with the risk of breast cancer, but it sounded small. The cardiovascular risk was not clear. She thought that she was too young for cardiovascular problems. The benefits were quite certain and they seemed to outweigh the risks so, like many other women, she was happy to continue HRT.

The example highlights many hidden risk perception and innumeracy issues. First, cognitive psychology experts would note the fact that benefits are conveyed as certain but that the harms are discussed as "risks." Many people process the term "benefits" at a subconscious level as certainty. By contrast, the term "risk" connotes uncertainty.[9] Thus, many physicians may inadvertently be presenting patients with certain benefit and uncertain harm. Second, there are important numeric evidence communication problems in this hypothetical case, as follows:

1. Relative risk estimates are not helpful for communicating the chance of experiencing an event (either benefit or a harm).[9,29] In our scenario, relative risk estimates cannot accurately depict the chance of experiencing menopausal symptom relief, fracture, or other benefits and harms.
2. Benefits related to menopausal symptoms and fracture are presented in relative terms, whereas harms (breast cancer) are described in frequencies. This approach is very likely to produce an asymmetric bias toward showing greater benefits and diminishing the importance of the harms.[9]

A Possible Solution

1. Dr. X should consider communicating the chance of a reduction in hot flashes in other terms. A 50% reduction in hot flashes associated with HRT might mean that 5

out of 10 women will be prevented from having any hot flashes, with no effect on others. However, it might also mean that the number of hot flashes will be halved in all women. There is also an intermediate scenario, with some women completely relieved, some having reductions, and some having no relief at all. The lack of clarity for this end point might need to be addressed by scientists and physicians together.

2. Dr. X should consider communicating the chance of the reduction in fractures in frequency terms as well. Example: 7–20 additional fractures will be prevented per 10,000 women taking HRT for 10 years.[27]

3. Although absolute risk terms are also appropriate, breast cancer harm can be communicated in terms similar to the above for consistency. Example: 18–57 additional breast cancers will occur in 10,000 women taking HRT for 10 years.[27]

4. The chance of stroke harm can also be put in similar terms. Example: about 92 additional strokes will occur in 10,000 women taking HRT for 10 years.[20]

More outcome data can be added in a similar symmetric manner.

If women were to learn that the benefits are not certain and a large number of women will not benefit or will get limited benefit from HRT but will still have a chance of having a harmful event, then they might make better decisions on the use of HRT. It is not helpful to communicate only the 50% reduction in menopausal symptoms such as hot flashes, as it might sound like a certain benefit applying to all who take HRT. Additionally, having frequency data on other relatively rare benefits and harms might help women to make their own judgments on how likely it is that will happen to them.