Walter C. Alvarez, who went above and beyond instead of donning his slippers at retirement, Helen spurred AMWA members to be Health Literacy Heroes, always remembering that our words matter.

Notes

aHelen Osborne holds a Master of Education and is an Occupational Therapist Registered/Licensed. She is President of Health Literacy Consulting (www.healthliteracy.com) and producer and host of the podcast series “Health Literacy Out Loud” (www.healthliteracyoutloud.com). She can be reached at: helen@healthliteracy.com


Warm thanks go to Marjorie Winters for her essential collaboration.

Liz Kuney, MS CCRP, is a Senior Medical Writer at BioTelemetry Research. She works from her home office in Syracuse, NY.

Author contact: lizkuney@gmail.com

John P. McGovern Award

Communicating Benefit and Harm: Avoiding the NNE (Numbers Needed to Exaggerate)

By Steven Woloshin, MD, MS; Lisa M. Schwartz, MD, MS; Emma Woloshin

The award, named in honor of John P. McGovern, is presented to a member or non-member of AMWA to recognize a preeminent contribution to any of the various modes of medical communication. The McGovern Award is presented during AMWA’s Medical Writing & Communication Conference. The 2017 recipients are Steven Woloshin, MD, MS, and Lisa M. Schwartz, MD, MS.

The McGovern Award is especially meaningful because it comes from a group of writers committed to fairly and effectively communicating complex medical information to the people who need it. We know how hard that can be.

Our work has focused on helping people make good decisions.1 What we have learned over the past 20 years boils down to this simple model: people need the facts and clarity about their values to make good decisions. This model falls apart without the facts. Unfortunately, there are a lot of bad facts.

Consider a Sloan-Kettering Cancer Center ad that says, “The early warning signs of colon cancer: You feel great. You have a healthy appetite. You’re only 50.” The message is meant to persuade people to get screened—not to inform them. This kind of scary message undermines peoples’ resilience. Don’t trust how you feel. If you feel healthy, you are probably sick. Yet most people who feel great, have a healthy appetite, and are 50 do not have colon cancer and will not get it. This ad—like many exaggerated messages—uses hype to generate fear to encourage some action.

Exaggerated messages also use hype to generate hope. For example, the president of one of the country’s major cancer centers told CNBC, “It’s actually plausible that in 10 years we will have curative therapies for most if not all human cancers.”2 That was in 2015. While it hasn’t happened yet, maybe it will in the next 8 years. But that seems pretty unlikely. Hopefully the cancer center president is more accurate than U.S. News and World Report. It predicted the end of heart disease back in 2003; as we all know, it’s still here.

Many exaggerated messages use both hope and fear to promote their product. Mount Sinai Medical Center advertises “An aneurysm is a death sentence. We have the power to grant you a pardon.” But in fact, most aneurysms are not a death sentence; the typical aneurysm found by screening is small and unlikely to cause problems, let alone death. Unfortunately, a small number of people who go for surgery will be hurt, and some may die, so that is hardly a pardon.

Given this state of exaggeration, consumers, clinicians, journalists, and medical writers need to develop a healthy skepticism to help their readers see through and push back
against hype. This includes a commitment to clearer, more complete, and transparent communication. Many messages like the foregoing examples could be improved by presenting numbers. But the numbers need to be presented fairly. All too often, numbers are used to amplify exaggeration—sometimes deliberately, but often inadvertently, because writers and editors may not understand basic statistics or the principles of how to communicate them.

Most people have heard of the NNT—the number needed to treat. Our talk is about another concept, NNE—numbers needed to exaggerate. That is, how common statistics, when misunderstood and misused (or intentionally manipulated), exaggerate the magnitude of a difference to make treatments seem better or safer than they really are. Unlike many problems in medicine, this is pretty easy to fix. With just a little background, you can see through misleading statistics and help your readers see through them. In this paper, we review 3 NNEs: changes in risk, odds ratios, and survival statistics and cancer screening.

**Changes in Risk**

Figure 1 is a medical journal advertisement for Evista (raloxifene), an osteoporosis drug, which claims that “Evista significantly reduces clinical vertebral fracture risk at one year—68% reduction risk versus placebo.”

While it’s easy to criticize drug ads, you can actually learn a lot from them. This one does some things right. It is clear about the outcome that the drug affects: clinical vertebral fractures, or broken bones in the spine. Some vertebral fractures are silent—a person doesn’t feel anything and they only know about the fractures because of an x-ray (or might notice a gradual loss in height). Clinical fractures are different. They hurt. Preventing painful fractures is important. The ad is also clear about the time frame: this is the benefit of the drug at 1 year.

The ad, however, is not clear about the most prominent element—the 68%. That’s a big number. It invites people to assume the drug works very well. How big is the benefit of Evista? You can’t tell from the ad. The impressive “68% reduction” is meaningless unless you know 68% lower than what. It’s like a sale. Imagine a store advertises a 68%-off sale on selected items. Would you drive hours to go to the store? Yes—if Ferraris were on sale. Not if packs of gum were on sale. A sale on a Ferrari saves you tens of thousands of dollars. A sale on a pack of gum saves you pennies. To decide how good a sale is, you need to know the regular price of the selected items. 68% of what matters.

The math behind the benefit of the drug is the same as a sale. The difference is the units. For a sale, you save dollars. For example, if the regular price is $100 and the sales price is $90, how much do you save? The savings is just the regular price minus the sales price: $100 − $90 = $10. For a drug, you save percentage points of risk. Let’s take a look at the science behind the Evista ad. In the trial, about 4500 postmenopausal women with osteoporosis at high risk for fractures were randomized to receive either Evista or a placebo. The primary outcome was the percentage of women who had a clinical vertebral fracture at 1 year. In the placebo group, 19 out of 2290 women had a fracture, which divides out to be 0.83%. This is called the absolute risk in control group (or base rate or event rate). That’s the regular price. In the Evista group, 6 out of 2259 women, or 0.27%, had a fracture. That’s the sales price. How much do you save? The savings is the regular price minus the sales price: 0.82% − 0.27% = 0.56%. The savings is called the absolute risk reduction. Evista lowered the risk of a vertebral fracture by 0.56% percentage points compared to placebo over 1 year. In other words, if a thousand women with osteoporosis at high risk for a fracture took Evista instead of placebo for 1 year, about 6 fewer would have a vertebral fracture.

The same data is more typically presented as a relative, not absolute, difference. The relative risk is a ratio: the absolute risk in the drug group divided by the absolute risk in the control group. The relative risk for Evista is 0.27% ÷ 0.83% = 0.32. The risk of vertebral fracture for women taking Evista was 0.32 times that of women taking placebo. Because of the clunky language, we usually talk about “percent lower” (the sale). The “percent lower” or “percent off”
(like a sale) rather than “times the risk” (have you ever heard a sales clerk say, this tie is on sale, it’s only 90% of the regular price?). equals 1 minus the relative risk. For Evista, that’s 1−0.32 = 0.68. That’s why the ad says “68% lower.” It’s a “68% off sale” on your vertebral fracture risk. Evista lowered the risk of vertebral fracture at 1 year by 68% compared to placebo: 0.27% vs 0.83%. The analogy between a sale and the medical risk reduction jargon is shown in the Box.

The foregoing example shows that the same data can be expressed in many different ways. Different formats have different psychological impact, a phenomenon called framing. It is a well-described finding that the relative risk reduction feels much more impressive than corresponding absolute risk reduction, especially when the outcome is uncommon, as illustrated in Table 1. The Evista ad agency knew what they were doing. Compare 68% lower versus 0.56% lower (Figure 1). Which will sell more Evista? It is now widely accepted that showing the absolute risks for both groups is the fairest way to present the data.

Another commonly used statistic is the NNT—the number needed to treat to prevent 1 outcome. It is simply the reciprocal of the absolute risk reduction. For Evista, the absolute risk reduction was 0.56% (ie, about 6 fewer clinical vertebral fractures over 1 year per 1,000 women). The NNT is just 1 ÷ 0.56%, which works out to 178. That means that 178 women have to take Evista for 1 year to prevent 1 vertebral fracture. Table 1 also includes the NNT to illustrate that as events become less common, the NNT increases.

Unfortunately, reporters often do not follow the recommendation to present absolute risks. Not long ago, all top 10 circulation US newspapers and Time magazine’s lead story covered the World Health Organization’s report linking consumption of processed meats to colon cancer. USA Today’s headline (“Hot dogs, bacon, processed meats linked to cancer”) and results reporting (“increases the risk of colorectal cancer by 18%”) were typical. Just like the Evista ad: giving the sale without the regular price (18% off what?). To their credit, the USA Today story translated the type and amount of processed meat that raised risk (e.g., 1 ¼ hot dogs a day), a statistic we call—with apologies to the NNT—the “number needed to eat.” But the story did not give the basic information needed to decide whether the risk of eating processed meats was big or small. The story should have reported the absolute risks, such as, “meat consumption increases the risk of colon cancer over your lifetime from 5% to 6%.”

What about harms? While some drugs have benefits, all have harms. Let’s return to the Evista ad. Prescription drug ads are required to present side effects. Accordingly, the bottom of the ad reads: “Evista is associated with an increased risk of venous thromboembolic events (potentially fatal blood clots).” The presentation format is unfair because it magnifies the benefit (68% reduction in big font without the small absolute risk reduction) but minimizes the harm (small font, no numbers, buried in a long list readers skip over).

To be fair, benefits and harms should be similarly prominent and quantified. To do this, we read the medical journal article that was the basis for the Evista ad and were surprised to see that the trial lasted 3 years. Trial 1 shows why: the risk reduction was lower (ie, the “sale” got less impressive). The ad should have read “Evista reduced clinical vertebral fractures by 41% at 3 years,” not “68% at 1 year.” Perhaps the advertising agency wanted to highlight the more impressive 1-year results.

If the ad agency quantified the harms the same way as the benefits, the ad would read “Evista increases thromboembolic events 210% versus placebo.” Many women would probably think twice about taking Evista. But wait a minute. Remember the advice: present absolute risks. This is just as true for harm as benefit. You can’t know what “210% more”
means unless you know 210% more than what. The risk of thromboembolic events in the placebo group—the “regular price”—was low: 0.35%. Consequently, Evista increased the risk from 0.35% to 1.1%, which is a much less dramatic difference.

To make good decisions about prescription drugs, people need balanced benefit and harm data. One morning over breakfast, the nutrition facts box—the consistent, structured data table required on all food packaging—inspired us. If we can do that for Cocoa Krispies, why can’t we do that for Evista? Not give the ingredients, but present the data on how well it works. Thus, we created the Drug Facts Box: a 1-page summary of benefit and harm data (absolute numbers) for each indication of a drug (Figure 2). A series of studies demonstrates that most consumers understand the Drug Facts Box and that it improves decision-making.11-15 In a nationally representative randomized trial (n = 231), 68% of people randomized to see direct-to-consumer advertisements with drug boxes chose the objectively better of 2 heartburn drugs, compared with 31% of people seeing standard advertisements.11 Based on our research, the FDA’s Risk Communication Advisory Committee voted unanimously in a nonbinding recommendation that the FDA “should adopt the Drug Facts Box format as its standard for communicating essential information about pharmaceuticals” (in the drug label for prescribers, advertisements, and other consumer materials). After The New York Times reported on our Advisory Committee presentation, 2 senators drafted a bill that was incorporated into the Affordable Care Act, Section 3507, encouraging the FDA to adopt the Drug Facts Box format. Although the FDA replicated our research findings, they decided not to implement Drug Facts Boxes. Fortunately, Consumers Reports,16 The New York Times,17 and National Public Radio18 have featured drug boxes. Recently, the UK Academy of Medical Sciences issued a report identifying the Drug Facts Box as an exemplar for the European Union and the European Medicines Agency to draw on to improve the content and accessibility of drug information for the public.19 So, we’re hopeful that at least Europe may adopt Drug Facts Boxes.

**Odds Ratios**

Our interest in odds ratios began when we heard Ted Koppel say on Nightline: “Last night we told you how the town of Jasper, Texas, is coming to terms with being the place where a black man was dragged to his death behind a truck by an avowed racist. Tonight, we’re going to focus on a group of men and women who are, almost by definition, humanitarians, who would be shocked to learn that what they do routinely fits quite easily into the category of racist behavior.” Who were these racists? We were shocked when Koppel’s guest, the Surgeon General of the United States, said these racists were American physicians. This story was big news, reported by every major media outlet.

We took a close look at the research article behind the news, a New England Journal of Medicine article titled “The Effect of Race and Sex on Physicians’ Recommendations for Cardiac Catheterization.”20 The article stated that women were referred for catheterization less often than men (odds ratio = 0.6) and that black people were referred less often than white people (odds ratio = 0.6). The typical news story reported “Blacks and women with chest pain are 40% less likely than whites or men to be referred by physicians for cardiac catheterizations.” Where does the 40% come from? It’s the “percent lower” format, in this case, calculated as 1 − 0.6 = 0.4, or 40%. To know what this means, you need to know that the percent of white people or men referred for catheterization was 91% (reported in Table 4 in the article, not in the abstract). To calculate the percent of black people or women referred, you could take “40% off” of 91%, which is 55%. But Table 4 reported that 85% of black people or women were referred. What happened? Odds and odds ratios are not the same as risks and relative risks: 40% lower odds cannot be assumed to be a 40% lower risk.

---

**Figure 2.** Drug Facts Box for Lunesta (eszopiclone).
Risk and odds are different ways to measure chance. A risk is a ratio of occurrence to the whole—a proportion—that ranges from zero to 1. The risk of heads in a flip of a coin is 1 over 2, or 50%. The risk of getting a “1” in a roll of a die is 1 over 6, or 17%. Odds are a ratio of occurrence to non-occurrence. The odds of getting heads in a flip of a coin is 1 over 1, or 1. The odds of getting a “1” in a roll of a die is 1 over 5, or 20%.

In the *New England Journal of Medicine* article, which stimulated the extensive news coverage, including Koppel’s show, 85% of black people and 91% of white people were referred for catheterization. The corresponding odds are 5.5 and 9.6. The odds ratio is merely a ratio of odds: 5.5/9.6 = 0.6. The relative risk, however, is 85%/91% = 0.91. It is highly doubtful the finding “blacks were 8% less likely to be referred” would have triggered a special episode of Nightline. One way to correctly report the odds ratio finding is “the odds that black people or women with chest pain were referred for cardiac catheterization are 40% lower than those of white people or men.” Most readers are unlikely to recognize the subtle difference between odds and risks, so this rewrite is unlikely to help. The simplest solution is to present the absolute risks in each group. The study shows 85% of black people or women are referred compared with 91% of white people or men.

But even this statement is not right. This study involved 722 physicians attending a primary care meeting. Each physician saw a randomly chosen video of an actor pretending to be a patient with chest pain and was asked “would you refer this patient for cardiac catheterization?” The results are hypothetical answers about actors. The main results looked at white versus black people and women versus men. The results further broken down by actors were: 91% of white men were referred, 91% of black men, 91% of white women, and 79% of black women. How can this be racism or sexism if white and black men are referred at the same rate and white men and women are referred at the same rate? In fact, only 1 of the 8 actors was referred at a markedly different rate—the older black woman. This raises more questions; was it about how she read the script or something else?

The follow-up to this story is that the *New England Journal of Medicine* published our critique as a full article, with an editorial note taking responsibility for the over-interpretation. Many newspapers and the Associated Press printed corrections. Ted Koppel, however, according to the show’s producer, stood by his story and did not issue a correction. The take-home message is that odds ratios should be translated into relative risks. As a rule of thumb, odds ratios will only differ importantly from relative risks when the base rate is 20% or higher. But the most important message is to find, and report, absolute risks.

### Survival Statistics and Cancer Screening

If there were an academy award for the most misused statistic, the winner would be survival statistics for cancer screening. Sadly, countless examples illustrate how survival statistics are misused. A few years ago, we published a critique of a prototypical example, the Komen Foundation’s “breast cancer awareness month” campaign ad, which says, “Early detection saves lives. The 5-year survival rate for breast cancer when caught early is 98%. When it’s not? 23%” (Figure 3). The enormous difference—98% versus 23%—makes you feel it would be crazy not to have a mammogram. The idea that early detection saves lives is intuitively obvious. The logic is the more people screened, the more people diagnosed with early cancer, and early cancers have higher survival rates than later-stage cancers; therefore, fewer people are dying from the disease. This is a giant leap of faith.

To understand why, the first step is to understand the difference between survival and mortality. Survival is the number of people alive some number of years after diagnosis divided by the number of cancer patients. If 980 out of 1,000 cancer patients are alive 5 years after diagnosis, 5-year survival is 98%. Mortality is the number of people who died some...
The 5-year survival is 20%: 200 women alive 5 years after diagnosis divided by 1,000 with breast cancer. Imagine exactly the same city, except women undergo screening. But now an additional 4,000 women are diagnosed because a mammogram found a cancer—none of which were destined to progress. That means all 4,000 will still be alive 5 years later. What would this do to 5-year survival? The number of women alive 5 years after diagnosis is 4,200 (4,000 found by screening + 200) out of 5,000: Five-year survival increased from 20% to 84%, but the same number of women died (800). This scenario is “overdiagnosis”—the detection of cancer that would never progress to cause symptoms or death. Overdiagnosis does not mean screening cannot work or that screen-detected cancers are never dangerous. It just means that comparing survival for screened and unscreened populations is inherently misleading.

Figure 4 shows 2 reasons for the confusion. First, imagine 100 women who do not undergo screening and are diagnosed at age 67 when they feel a breast lump (Figure 4A). The black line represents the survival time, or how long they live past the time of diagnosis. Assume the women die of breast cancer at age 70. For this unscreened group, the 5-year survival is 0%: no woman survived 5 years past diagnosis. The second arrow shows what happens to exactly the same women with screening. Because mammograms can pick up cancers earlier, the women are diagnosed at age 63. Again, assume the women still died at age 70, 5-year survival would be 100% because all women lived more than 5 years past diagnosis. Yet none lived even a second longer. This is called lead time bias, and it is a mathematical certainty whenever you diagnose disease earlier.

Dr Barnett Kramer, Director of the National Cancer Institutes’ Division of Cancer Prevention, explains lead time bias with an analogy to the old Bullwinkle cartoon. In a recurring segment, Snidely Whiplash ties Nell Fenwick to the railroad tracks to extort money from her family. She will die when the train arrives. Kramer says that lead time bias is like giving Nell binoculars. She’ll see the train when it’s further away, but it will hit her at exactly the same moment. She just lived longer with the diagnosis of “train.” Lead time bias does not mean that screening cannot work. It just means that comparing survival for screened and unscreened populations is inherently misleading.

Figure 4B shows the second reason why comparing survival for screened and unscreened groups is misleading. Imagine a large city where no one undergoes screening. Say there are 1,000 women with breast cancers diagnosed because they felt a lump. After 5 years, 200 women are alive and 800 are dead. The 5-year survival is the proportion of women alive 5 years later divided by the entire population (not just cancer patients). If 10 out of 1,000 people died 5 years after the study, mortality is 1%. Mortality typically comes from a randomized trial.

In the context of screening, impressive 5-year survival statistics do not necessarily mean lower mortality—that is, that any lives have been saved. Survival statistics are about what happens after the time of diagnosis, and screening changes when a cancer is diagnosed. Without screening, breast cancers are only detected when a tumor is big enough to feel. With screening, cancers can be detected much earlier, often when the tumor is much too small to feel. Five-year survival is the proportion of women alive 5 years after the cancer is diagnosed. Even if the cancer was untreated, 5-year survival improves with screening just because the clock starts earlier. It is, of course, possible that earlier treatment may help, but that is not the issue here. The issue is that simply comparing survival between screened and unscreened women is inherently misleading.

In the context of screening, impressive 5-year survival statistics do not necessarily mean lower mortality—that is, that any lives have been saved. Survival statistics are about what happens after the time of diagnosis, and screening changes when a cancer is diagnosed. Without screening, breast cancers are only detected when a tumor is big enough to feel. With screening, cancers can be detected much earlier, often when the tumor is much too small to feel. Five-year survival is the proportion of women alive 5 years after the cancer is diagnosed. Even if the cancer was untreated, 5-year survival improves with screening just because the clock starts earlier. It is, of course, possible that earlier treatment may help, but that is not the issue here. The issue is that simply comparing survival between screened and unscreened women is inherently misleading.

Figure 4 shows 2 reasons for the confusion. First, imagine 100 women who do not undergo screening and are diagnosed at age 67 when they feel a breast lump (Figure 4A). The black line represents the survival time, or how long they live past the time of diagnosis. Assume the women die of breast cancer at age 70. For this unscreened group, the 5-year survival is 0%: no woman survived 5 years past diagnosis. The second arrow shows what happens to exactly the same women with screening. Because mammograms can pick up cancers earlier, the women are diagnosed at age 63. Again, assume the women still died at age 70, 5-year survival would be 100% because all women lived more than 5 years past diagnosis. Yet none lived even a second longer. This is called lead time bias, and it is a mathematical certainty whenever you diagnose disease earlier.

Dr Barnett Kramer, Director of the National Cancer Institutes’ Division of Cancer Prevention, explains lead time bias with an analogy to the old Bullwinkle cartoon. In a recurring segment, Snidely Whiplash ties Nell Fenwick to the railroad tracks to extort money from her family. She will die when the train arrives. Kramer says that lead time bias is like giving Nell binoculars. She’ll see the train when it’s further away, but it will hit her at exactly the same moment. She just lived longer with the diagnosis of “train.” Lead time bias does not mean that screening cannot work. It just means that comparing survival for screened and unscreened populations is inherently misleading.

Figure 4B shows the second reason why comparing survival for screened and unscreened groups is misleading. Imagine a large city where no one undergoes screening. Say there are 1,000 women with breast cancers diagnosed because they felt a lump. After 5 years, 200 women are alive and 800 are dead. The 5-year survival is 20%: 200 women alive 5 years after diagnosis divided by 1,000 with breast cancer. Imagine exactly the same city, except women undergo screening. But now an additional 4,000 women are diagnosed because a mammogram found a cancer—none of which were destined to progress. That means all 4,000 will still be alive 5 years later. What would this do to 5-year survival? The number of women alive 5 years after diagnosis is 4,200 (4,000 found by screening + 200) out of 5,000: Five-year survival increased from 20% to 84%, but the same number of women died (800). This scenario is “overdiagnosis”—the detection of cancer that would never progress to cause symptoms or death. Overdiagnosis does not mean screening cannot work or that screen-detected cancers are never dangerous. It just means that comparing survival for screened and unscreened populations is inherently misleading.
Because lead time and overdiagnosis biases cannot be disentangled, experts have long pointed out that, in the context of screening, survival statistics are meaningless. What is meaningful—and what matters a lot—is whether screening changes how many people die from the cancer. The benefit of breast cancer screening looks more like this. Imagine 1,000 50-year old women over 10 years. Five will die from breast cancer without screening versus 4 with screening. So, 1 out of 1,000 women avoids a breast cancer death because of screening.

**Conclusion**

Medical writers have a very important but very hard job. Effectively communicating complex medical information to the public is crucial. An uninformed public is especially vulnerable. Effectively communicating complex medical information to the public is crucial. An uninformed public is especially vulnerable. An uninformed public is especially vulnerable.

**References**


**Conflict of Interest Statement:** Drs Schwartz and Woloshin have served as medical experts in testosterone litigation and were the cofounders of Informulary, Inc, a company that provided data about the benefits and harms of prescription drugs, which ceased operations in December 2016. Ms Woloshin has no conflicts to report.

**Corresponding author:** Lisa M. Schwartz, MD, MS, Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756. Email: lisa.schwartz@dartmouth.edu