
**Los modelos de previsión reducen la incertidumbre, no la anulan**

**Elmore / Ventura**

# Seeing through the hype: Garbage! When the news is not fit to print

Steven Woloshin and Lisa M. Schwartz

Sometimes the medical news may not be fit to print: the research is so preliminary or so inherently weak that reporting it would more likely mislead than inform the public. For example, stories about a new miracle diet, or cancer breakthrough –stories which often arise out of abstracts presented at scientific meetings that have not undergone peer review, or uncontrolled human studies. Nevertheless, journalists often feel pressure to report on such studies.

What should journalists do when the news is not fit to print? Ideally they would not report it. But if they have to report it, they should always include strong cautions to alert readers to questions about the validity, meaning or generalizability of the research. In this essay, we will use real examples to review cautions about research that may sound exciting but is very preliminary or inherently weak.

**"Raloxifene may decrease the risk of endometrial cancer in post-menopausal women" Meeting abstract, American Society of Clinical Oncology 1998 meeting**

When this abstract was presented at a plenary session at the 1998 ASCO meeting, several news outlets covered the story, including the *Wall Street Journal*.[1] The abstract reported on what seemed to be a major advance: a relatively new medication which behaved differently than others in its class which increased the risk of endometrial (uterine cancer).

The investigators presented results from a randomized trial of raloxifene, a selective estrogen receptor modulator, or placebo which included 7704 postmenopausal women (mean age 66.5 years) with osteoporosis (based on hip or

spine bone density at least 2 standard deviations below normal or a history of vertebral fractures). The study was "designed to test the hypothesis that women assigned to receive raloxifene will have a lower risk of fractures than women assigned to placebo". But the main result reported in the ASCO abstract was not about the primary outcome –osteoporosis– but about endometrial cancer. Of course, endometrial cancer is not the same as osteoporotic fractures.

While it is perfectly legitimate for a study to report on multiple outcomes, these outcomes need to be specified in advance. Otherwise, surprise findings –which may reflect chance alone may be over interpreted. That is, taken as strong evidence for a treatment effect rather than a hypothesis generating finding which needs confirmation in a subsequent trial. For example, if 20 random outcomes are assessed after a trial is completed, 1 will be statistically significant –have a p value less than 0.05– just by chance. While endometrial cancer was a pre-specified study outcome, the researchers' hypothesis was the opposite: they were concerned it might increase cancer, a harm associated with raloxifene's rival, the drug tamoxifen. In fact, according to the trial protocol, assessing the drug's effect on endometrial cancer was designated as a safety issue, not as a potential benefit.[2]

Contrary to their hypothesis, they found that raloxifene *decreased* endometrial cancer: since this decrease in endometrial cancer was not a pre-specified hypothesis, the finding should be considered a "surprise" and interpreted cautiously.

How much did raloxifene decrease the risk of endometrial cancer? According to the abstract, "compared with the rate in the placebo group, the overall relative risk of endometrial cancer is 0.38 (p = 0.232). If two cases diagnosed within one month of randomization are excluded, the estimate of relative risk is 0.13 (p = 0.045)". While one may question the legitimacy of excluding data (particularly when doing so generates a desirable finding), the rationale for exclusion may be reasonable –the drug is unlikely to have had such a rapid effect, so the two cancers were probably present at the time of randomization. It would be reassuring to know that such exclusions were written into the study protocol, rather than excluded after the data were analyzed (we have not been able to ascertain which was the case in this trial). Another theoretical concern about this finding is whether the scrutiny was similar for both the raloxifene and placebo groups. If not, the investigators will have introduced bias. For example, if researchers looked harder for endometrial cancers in women in the raloxifene group compared to the placebo group (and excluded any found) this would bias the study in favor of the drug because preexisting cancers would still count against placebo. Fortunately, based on the protocol, the level of scrutiny was appeared to be the same in both groups in this trial.

Assuming the exclusions were legitimate, the fact that removing two cases had such a dramatic effect highlights a second concern: the results are very unstable. The magnitude of the risk reduction increased from 62% (relative risk = 0.38 ) to 87% (relative risk = 0.13) and the p-value changed from *not* statistically significant (p = 0.232) to statistically significant (p = 0.045). This instability reflects the preliminary nature of the report. The study was not over –only a fraction of the final data was collected. While the researchers did not report the number of cancers found in either the raloxifene or placebo groups, the number must have been very small. Waiting a little longer to accrue more data, a few more cases might flip the findings back.

In fact, waiting for more data did reverse the finding. When final study results were reported in *JAMA,*[3] raloxifene no longer had any effect on uterine cancer (Table 1).

Not only were the findings different, the same authors made a big change to their message. In the scientific meeting abstract, they wrote that the "Raloxifene *may decrease* the risk of endometrial cancer in post-menopausal women". In the *JAMA* article published 1 year later, they wrote "Raloxifene *did not increase* the risk of endometrial cancer". The *JAMA* message reflects their original hypothesis that endometrial cancer was a safety concern.

Here is how we would have rewritten the findings presented at the scientific meeting: *"There was a trend toward a lower rate of uterine cancer*

**Table 1.**

| | Relative Risk (Raloxifene vs Placebo) | P-Value |
|---|---|---|
| 1998 ASCO Meeting | | |
| All data | 0.38 | 0.232 |
| Exclude 2 cases* | 0.13 | 0.045 |
| 1999 JAMA | 0.80 | 0.67 |

*Cases which ocurred within 1 month of randomization.

*but it may be due to chance and it is too early in the study to say. This was not the main outcome being studied. We are not very confident in these results".*

This example highlights the fundamental problem with early results –they turn out not to be true (as in this case) or they may change substantially. When they report on scientific meeting presentations, journalists should raise a red flag for their readers. Our suggestion for this cautionary note is "These preliminary findings may change because the study has not been independently vetted through peer review and all the data are not yet in". (Note: this caution and the ones that follow are summarized in the tip sheet: *How to highlight study cautions.* See Apéndice, p. 80.)

### "Drug advances bring new hope to cancer battle –New treatments block 'switches' that turn cells malignant" *Wall Street Journal*

Other major U.S newspapers echoed the excitement of the *Wall Street Journal* headline: "Drug shrinks lung tumors" *Washington Post,* "Major step in cancer fight" *Houston Chronicle,* "Pill shows significant results in battling advanced lung cancer" *The Milwaukee Journal Sentinel*. Lung cancer is a terrible disease, one for which we do not have very effective treatments, so a real breakthrough would be wonderful news. Is this new drug really a breakthrough?

To understand whether this enthusiasm is warranted involves looking at the science behind the headlines.[4] The study followed 216 patients with advanced lung cancer who were all given the new treatment –*Iressa*. Unfortunately, there was no control group. Without a control group, it is extremely difficult to learn much about how well the treatment works. It is possible that equivalent patients who did not get the drug would have done worse (meaning *Iressa* helped). But it is also possible that the patients would have done no better or even worse (meaning *Iressa* caused harm). Here's the first red flag for journalists and readers –"Because everyone took Iressa, it is extremely hard to know if *Iressa* had anything to do with the outcome".

Even if this study were a randomized trial, a second fundamental problem exists: the primary outcome was a surrogate measure –tumor shrinkage (by half or more). The study found that 10% of the 216 patients had tumor shrinkage.[5] It is a big leap of faith to assume that tumor shrinkage means less suffering or death from lung cancer. There are three reasons why this is a leap of faith. Tumor shrinkage may be followed by period rapid growth. Or the tumor may shrink in an unimportant area that does not affect a person's health. Finally, spread in rest of body may be much more important than tumor shrinkage. So again, readers need a cautionary note: "This study measured tumor shrinkage –an x-ray finding that patients do not directly experience. Be cautious about acting on these findings since changes in these kinds of measures don't reliably translate into people feeling better or living longer".

Despite these two fundamental limitations, the study received a lot of enthusiastic press. This initial press coverage illustrates the beginning of an unfortunately common cycle. The cycle begins with great news –typically with breathless excitement about a new technology. But terrible news quickly follows –when side effects start to emerge as more people take the drug. In the case of *Iressa,* this happened with reports of drug-related

deaths in Japan where the drug was already approved for the treatment of advanced small cell lung cancer. Here is an excerpt from the *Wall Street Journal's* story titled "AstraZeneca drug used to fight cancer is tied to 124 deaths": "Side effects from the cancer-fighting drug Iressa have resulted in 124 deaths in Japan, a government official here said, as a ministry panel set stricter guidelines for the drug's use. .... Early studies showed lung-cancer patients who hadn't been helped by other therapies recovered impressively after taking Iressa [the impressive recovery refers to the study above where only 10% of patients had tumor shrinkage], but the large number of severe side effects prompted AstraZeneca to issue a safety warning to Japanese doctors in October".

Despite the report of deaths in Japan, the U.S. Food and Drug Administration (FDA) approved *Iressa*.[6] In fact, this was their first accelerated approval –a new program to more rapidly approve drugs for dangerous diseases with limited treatment options– like lung cancer –based on early studies. Accelerated approval is typically conditional upon confirmation of the results in a randomized trial post-approval. The *Wall Street Journal* was extremely enthusiastic about this decision. Their editorial board wrote: "A rare victory at the FDA. When an FDA advisory panel convened Tuesday to consider AstraZeneca's application for the cancer drug Iressa, it was expected to send the company back for more data. But spurred on by powerful testimony from patients who would almost surely be dead without the drug, and over the apparent objections of hypercautious FDA staffers, the panel voted 11-3 to recommend Iressa for accelerated approval".

Hope was even higher when another study was published in *The New England Journal of Medicine*.[7] This study received hyperbolic coverage –particularly on the U.S. national news. NBC national news ran a segment "Scientists announce major breakthrough in treatment of lung cancer with *Iressa*". The segment featured the story of a young woman with children whose tumor melted away and quotes experts who say the drug will save thousands of lives. Surprisingly, the *The New England Journal of Medicine* article was reporting on 16 of the patients from the original *Iressa* study (where all of the patients received *Iressa*). The "new" study found that eight of the nine patients who responded to Iressa (experienced tumor shrinkage) had a specific genetic mutation while none of the seven patients who had not responded to *Iressa* lacked the mutation. But since all patients with the mutation received Iressa, there is (as described above) no way to know if the mutation predicted response to the drug. Unfortunately, subsequent work failed to confirm that the genetic mutation predicted response to this class of drugs.

Even worse, the phase III study (required by the FDA as part of the accelerated approval program) did not find any survival benefit from *Iressa*. In this randomized trial of 1,700 lung cancer patients, the *Iressa* group had a median survival of 5.6 months vs. 5.1 months in the placebo group.[8] The FDA then pulled the drug from the market, only allowing it for compassionate use.[9] Figure 1 summarizes the *Iressa* and the news cycle.

### "Major cancer breakthrough? New drug potential 'holy grail' for treatment of cancer" CBS Healthwatch, The Early Show

Excessive hope about cancer drugs did not end with Iressa story. A major breakthrough in 2009 started with a *The New England Journal of Medicine* article about "Parp inhibitors" (drugs which inhibit poly (ADP ribose) polymerase) in cancer patients who had BRCA mutations.[10] It is hard to exaggerate the exaggeration of the three major national television news networks.[11] In addition to the *CBS* story, *NBC Nightly* news reported "Now we turn to what some are calling the most important cancer treatment breakthrough in a decade", and *ABC* news "New hope: cancer treatment".

To distinguish hope from hype, we need to understand the science behind the headlines. This study measured what happened to 19 patients with BRCA1 or BRCA2 mutations with ovarian, breast or prostate cancer. After about 5 months of follow-up, 63% had either stable disease (stable tumor markers) or response (defined as 30% or more tumor shrinkage on x-ray). This study was a phase I study –an uncontrolled study

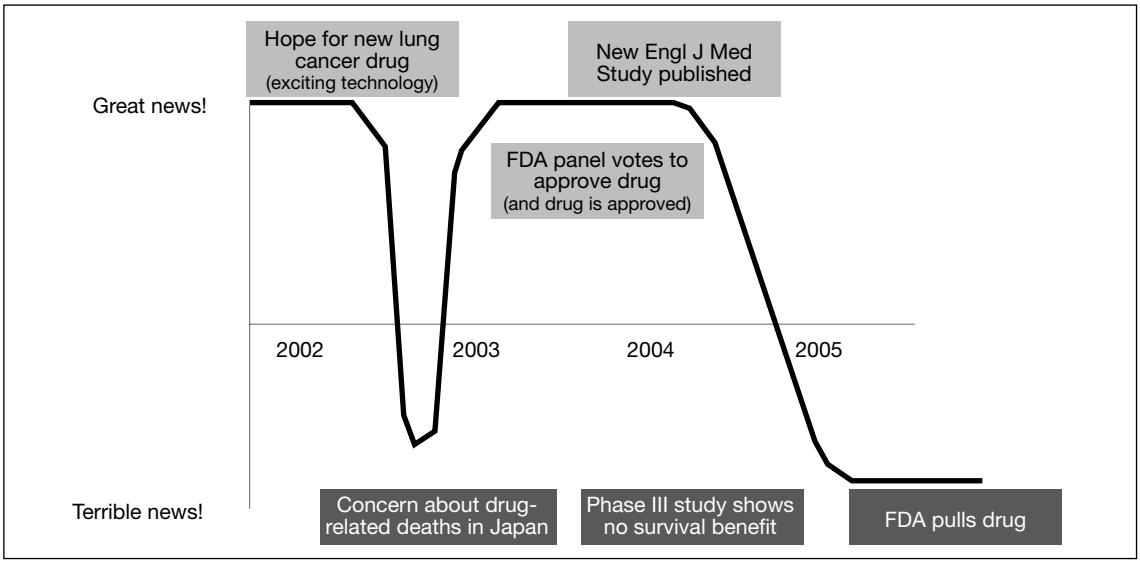


**Figure 1.**

using a surrogate outcome– just like the *Iressa* study. Publication in *The New England Journal of Medicine* seemed to trump the weakness of the science (which many journalists recognized). In fact, some asserted that since *The New England Journal of Medicine* typically does not publish this kind of uncontrolled study it must mean that this study was very important. Sadly, once again, the randomized trial did not show a difference in cancer death –and the drug company has abandoned pursuing approval.

The Parp inhibitor story reinforces the lessons of *Iressa* –be extremely cautious about uncontrolled studies and surrogate outcomes. But it also holds another important lesson: recognize pseudo-evidence. Publication in a medical journal –even *The New England Journal of Medicine*– does not guarantee the findings are true (or even important). We can all agree that giving false hope to sick patients is a real disservice.

### "Your friends may be as powerful as anticancer drugs in the fight against breast cancer" *Women's Day* (magazine), October 2010

"Do you get together with friends often? Here's an important reason to accept your pal's book club invitation: An active social life is not only good for your general health, but keeping up with your girlfriends may also reduce your risk of developing breast cancer. In a recent study, researchers at the University of Chicago report that lonely women may be at greater risk of breast cancer. The theory? Stress and anxiety caused by social isolation may have the power to increase the growth of tumors in the breast."

Is it really possible that all you need to do is spend some time with your friends to reduce your risk of breast cancer? This magazine story is based on a study published in the prestigious journal, *Proceedings of the National Academy of Sciences*.[12]

The story, however, wildly extrapolates findings from a study of 40 rats. Half the rats were randomized to live alone from 1 month of birth until death (the other lived in groups of five-female rats). Rats were just as likely to develop a breast tumor regardless of isolation; however isolated rats developed more and larger tumors.

Does this really mean that your friends are as powerful as "anticancer drugs"? Inbred rats, genetically altered so they are predisposed to develop breast cancer are not typical rats. And, of course, even typical rats are not like typical humans. Nor is total, lifelong isolation the same as refusing your pal's book club invitation or feeling lonely.

If we had to write about this study, here is what we would say: "News only a mutant rat could use (maybe). This study of mutant rats forced to live in total lifelong isolation has no direct meaning on cancer risk for humans or even ordinary rats. Don't get stressed out by this study of stress. And don't feel like you have to change your social behaviors. The level of social isolation in this rat study was far more extreme than any human being could ever experience."

While test-tube and animal studies can be fundamentally important, the problem is claiming imminent relevance to human health. In a systematic review of "high profile" animal studies, it took an average of 14 years to translate the animal research into human testing.[13] And only one-third of animal studies translated into successful interventions in randomized trials of humans. Moving from animals to humans is a slow and uncertain process.

When reporting on such research, extrapolate with caution. Do not tell people what to worry about –or do– based on very preliminary animal or lab science. We recommend cautioning readers that "It takes many years to learn if the findings of animal [lab] studies apply to people. Many promising animal [lab] studies fail to pan out in people".

Sometimes the news is not fit to print. Consider asking your editor whether you can skip it. If you have to report it, always include STRONG cautions. These will help readers avoid over interpreting the findings –and may even sway your editor against covering the study after all.

## References


1. Schwartz L, Woloshin S, Baczek L. Media coverage of scientific meetings: too much, too soon? JAMA. 2002;287:2859-63.
2. Comparison of raloxifene hydrochloride and placebo in the treatment of postmenopausal women with osteoporosis (MORE) trial protocol NCT00670319. Available at: http://clinicaltrials.gov/ct2/show/NCT-00670319www.clinicaltrials.gov
3. Cummings S, Eckert S, Krueger K. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. JAMA. 1999;281:2180-97.
4. Wolosshin S, Schwartz L. What's the rush? The dissemination and adoption of preliminary research results. J Natl Cancer Inst. 2006;98:372-3.
5. Pollack A. Drug's approval hints at flexibility in FDA process. New York Times. May 6, 2003;C; Column 5: 1.
6. United States Food and Drug Administration. Questions and answers on Iressa (gefitinib). 2005. (Accessed January 28, 2006.) Available at: http://www.fda.gov/cder/drug/infopage/iressa/iressaQ&A2005.htm
7. Lynch T, Bell D, Sordella R. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350:2129-39.
8. Pollack A. Lung-cancer drug shows unfavorable trial results. New York Times. December 20, 2004;C; Column 1: 2.
9. Pollack A. FDA restricts access to cancer drug, citing ineffectiveness. New York Times. June 18, 2005; C; Column 1: 2.
10. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361:123-34.
11. Woloshin S, Schwartz L, Kramer B. Promoting healthy skepticism in the news: helping journalists get it right. J Natl Cancer Inst. 2009;101:1596-9.
12. Hermes G, Delgado B, Tretiakova M. Social isolation dysregulates endocrine and behavioral stress while increasing malignant burden of spontaneous mammary tumors. PNAS. 2009;106:22393-8.
13. Hackam DG, Redelmeier DA. Translation of evidence from animals to humans. JAMA. 2006;296:1731-2.