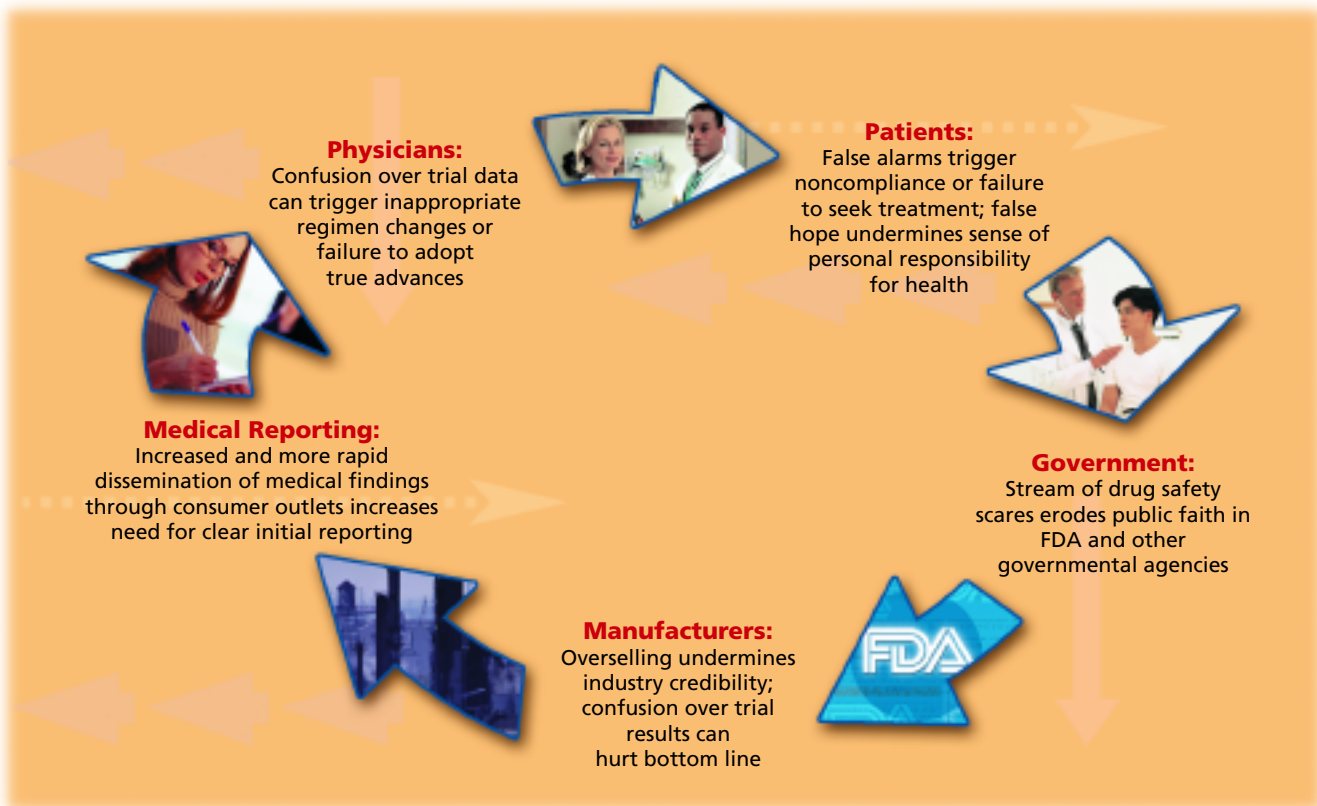


R. Brian Attig and Alison Clabaugh

# Clinical Trials and Statistical Tribulations



PHOTOGRAPHY: ARTVILLE; COMSTOCK; PHOTODISC ILLUSTRATION: PAULA BELICI

**Figure 1.** Cycle of influence in public discourse about medications.

.....

## The recent Avandia case and ensuing public discussion highlight the need for clearer statistical metrics.

.....

**C**onfusion over how to interpret the data that come out of many clinical trials is not a new phenomenon and sometimes seems endemic to the process. Most recently, concern over the safety of FDA approved medications has been fueled by a seemingly steady stream of examples, with GSK's diabetes drug Avandia among the most recently profiled. This confusion emanates from multiple

sources—manufacturers, academic journals, mass media reporting—and has far reaching negative consequences for much needed informed public discourse about new and existing medications.

A recent and widely publicized meta-analysis published in the June 14, 2007, issue of *The New England Journal of Medicine* by Nissen and Wolski<sup>1</sup> raised concerns about GSK's diabetes drug Avandia (rosiglitazone). The finding that received the most attention was a suggestion that it is linked to an increased risk of myocardial infarction (MI). Headlines and conclusions to emerge from this article almost universally ran along the lines of "43% more heart attacks with Avandia"—and this reporting was widely echoed across mass media publications, from *The New York Times*<sup>2</sup> and *Wall Street Journal*<sup>3</sup> to online outlets such

as MedicineNet.com,<sup>4</sup> as well as medical journals such as *The Lancet*.<sup>5</sup> Even the FDA entered the fray, calling for hearings and committees to examine Avandia's safety.

While we don't seek to minimize the concern expressed over this issue, what is largely missing from ongoing discussions about Avandia's safety is consideration of absolute risk levels—a topic perhaps less headline worthy, but nonetheless important as part of the overall clinical picture.

To understand the source of the concern and confusion about Avandia, one must first consider the primary statistical metric employed in the Nissen and Wolski meta-analysis, something called an odds ratio. Widely used in epidemiology, odds ratios are generally regarded as difficult to interpret and apply in medical practice. So, what exactly does an odds ratio tell us? A literal definition is “the odds of an event in the active treatment group (P1) divided by the odds of an event in the control group (P2)”<sup>6</sup> as expressed in the following formula:

$$\frac{P1/(1 - P1)}{P2/(1 - P2)}$$

Although often incorrectly taken as such, an odds ratio does not correspond to the probability of an event occurring or typically to the relative risk between groups (i.e., the ratio of two probabilities). A brief and simple illustrative example helps to clarify how an odds ratio can create confusion around research findings. As shown in Table 1, treated patients are 4.5 times more likely to show improvement than control patients (90% ÷ 20%). However, these same data can also be expressed as an odds ratio of 36 (by dividing 9 by .25).

Although the calculation of this odds ratio is clear enough, its interpretation is certainly less so on any intuitive level. Treated patients in this example are clearly not 36 times more likely to show improvement than control patients. Instead, the proper—albeit puzzling—interpretation is that the odds of treated patients showing improvement are 36 times greater compared to the odds of control patients. It seems quite fitting that the odds ratio has been called “a stranger to both physicians and gamblers, but a friend of many biostatisticians and epidemiologists.”<sup>7</sup>

One of the main figures from Nissen and Wolski's meta-analysis to create the widespread concern about Avandia is a reported odds ratio for MI of 1.43. As mentioned, this figure has been almost exclusively reported to mean a 43% increased risk of MI for those patients taking Avandia, a finding that on its own does sound alarming. When absolute levels of risk are compared between treatment and control groups, however, a somewhat more restrained picture emerges.

## Confusion about clinical trial results creates a negative feedback loop that affects all players in the health care system.

For reasons not relevant to this article, three separate subgroups of trials were compared in the Nissen and Wolski meta-analysis (small trials, DREAM, and ADOPT). As shown in Table 2, the absolute level of increased MI risk for Avandia patients in these groups ranged from 7/100th of 1% to slightly more than 4/10th of 1%, very small effects by any measure.

Amidst the recent Avandia debate, much discussion has focused on which trial results are more relevant: those in the Nissen and Wolski meta-analysis, those previously provided by GSK separately or those recently rushed into publication from the ongoing RECORD study. Attention has also been given to the issue of whether or not meta-analyses per se are appropriate and actionable in a clinical context. Largely absent in the ongoing public discussion about Avandia is a contextualization about the meaning of the 1.43 odds ratio, which ultimately is the primary source of recent concern.

If the Avandia meta-analysis was an isolated instance of confusion about trial results, this example would have limited relevance. But this case is one among many, which range from confusion about drug safety to unrealistic hopes for drug efficacy.

### Through the looking glass

Given its complicated scientific nature, along with various social and commercial forces at play, interpretation of pharmaceutical clinical trial data is rarely a straightforward matter.

For self-evident reasons, manufacturers seek to present

### Odds Ratio and Confusion: A Simple Example

Quota Group	Show Improvement	No Improvement	Odds Improvement: No Improvement	Odds Ratio
Treated Patients	90%	10%	9:1 (9)	36
Control Patients	20%	80%	1:4 (.25)	

**Table 1.** Odds ratio calculations for hypothetical product efficacy.

### Avandia Trial Results Up Close

Study	Avandia Group # of MI events/total # of patients (event rate)	Control Group	Odds Ratio	Absolute Risk Difference Avandia—Control
Small trials	44/10,285 (0.0043)	22/6106 (0.0036)	1.45	0.0007
DREAM	15/2635 (0.0057)	9/2634 (0.0034)	1.65	0.0023
ADOPT	27/1456 (0.0185)	41/2895 (0.0142)	1.33	0.0041
Overall (pooled using Peto method)			1.43	

**Table 2.** Avandia MI meta-analysis figures (adapted from Nissen and Wolski<sup>1</sup>).

data on their products in the most favorable light—one that typically involves the use of “relative risk” figures. Relative risk is defined as the probability of an event in the active treatment group divided by the probability of an event in the control group (and relative risk reduction is the inverse of this).<sup>6</sup>

It is a well-established psychological phenomenon that the framing effect of relative risk figures, divorced from a baseline of actual event rate, has a greater ability to persuade physicians to prescribe a medication than corresponding figures on absolute risk.<sup>8</sup> What sounds more impressive? New product ABC reduces MI by 50% compared to XYZ, or new product ABC reduces the incidence of MI from 2% to 1% compared to XYZ? Both claims can be made based on the same data (see Table 3).

At the other end of the spectrum, increasingly sophisticated data mining techniques have enabled researchers whose job it is to monitor drug safety to flag sometimes very slight signals of concern about possible medication side effects—ones that often do not show up initially during clinical trials. While this is obviously an important activity, one of the challenges this effort highlights is how best to assess the risk/reward profile for a particular medication. A recent example of this was reported in a *Wall Street Journal* article about a hypothesized link between statins and Lou Gehrig’s disease (ALS).<sup>9</sup>

A researcher at the World Health Organization’s Drug Monitoring Center noted that 40 of the 172 ALS patients in their database had been on statins, raising a preliminary although unproven connection. Even if this connection was to be established, how are we to evaluate the scenario? On one hand, ALS is deadly but rare, striking about five in 100,000 people. On the other hand, MI is a leading cause of death and statins have ostensibly helped millions of people reduce heart risk. We clearly go beyond the realm of strictly clinical data and into an area of subjective judgment when weighing such matters.

The need for clearly presented and clinically relevant information about medications is vitally important—and ever more so as patients increasingly avail themselves of information and proactively engage their health care providers in discussions. Confusion about clinical trial results creates a negative feed-

back loop that affects all players in the health care system (as detailed in Figure 1). In the Avandia case, the cycle of influence played out very quickly and with dramatic consequences—Avandia’s share of new prescriptions dropped from about 30% at the beginning of May 2007 (before publication of the Nissen and Wolski article) to less than 5% by the beginning of June 2007.<sup>10</sup> This has been a boon for Eli Lilly’s Actos (a direct competitor of Avandia with a similar mechanism of action), but the consequences are less clear for others.

Are doctors to assume this is a class effect (which could also implicate Actos)? Should patients be switched to older and “safer” diabetes medications, such as metformin and sulfonylureas, which may not control blood sugars as effectively and, thus, could increase patient risk in other ways? How many patients on Avandia have made their own decision to discontinue therapy? Despite pleas from various sides not to rush to judgment, that is in fact what happened on the frontlines of patient care in this case.

### Toward a grammar of risk assessment

A prevailing theme in much industry discussion over the past few years has been risk assessment: how best to quantify it and apply it to clinical practice. Relatively recent developments such as increased public availability of trial data (via ClinicalTrials.gov and other initiatives), along with increased activity in data mining these results, further raise the importance of developing more effective means to assess medication risk/reward.

Toward this end, one useful starting point in this effort could be the wider application of more intuitive reporting metrics for pharmaceutical trial results. After all, how are we to accurately assess risk if we don’t fully grasp the clinical meaning behind the numbers? A suggested best practice for any type of reporting (academic journals, manufacturer marketing materials, mass media reports) should include statements of:

- Absolute risk reduction: Reflects a baseline event rate for treatment and control groups, thus providing a framework for understanding the overall magnitude of risk (or benefit) as well as impact from the intervention.
- Relative risk reduction: Provides an easily understandable and direct means to compare the effect of an intervention on treatment and control groups.

Statements of both relative as well as absolute risk complement each other and facilitate the ability to effectively understand the impact and value of an intervention. For this reason, an odds ratio or relative risk measures should not be discussed in isolation from absolute risk (or benefit) levels.

## Risk assessment has been the theme of much discussion in industry over the past few years.

### Relative Risk vs. Absolute Risk Figures

#### Myocardial Infarction in Five Years

	Controls (c)	Active Treatment (a)	Relative Risk Reduction (Pc-Pa)/Pc	Absolute Risk Reduction Pc-Pa
<b>Patient Groups</b>				
Event rate (P)	2%	1%	50%	1%

**Table 3.** Despite being based on the same data, because of the way they are framed, relative risk figures sound more impressive than absolute risk figures.

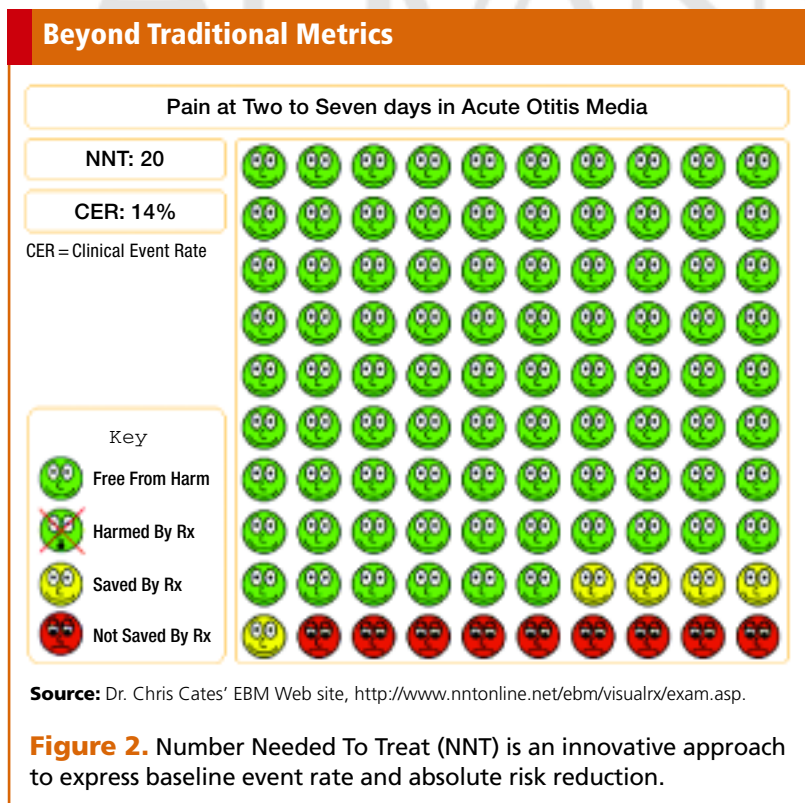
Some researchers have gone beyond traditional statements of risk reduction and proposed innovative approaches that seek to frame things in a more clinically salient context. One example of this is Number Needed To Treat (NNT). This metric was originally suggested by Laupacis and colleagues<sup>11</sup> and repre-

**It is in the best interest of all players in the health care system to promote and encourage efforts for clearer reporting of trial data.**

sents the number of people who need to be treated with an intervention in order that one of them receives benefit. It is the reciprocal of Absolute Risk Reduction (AAR) (e.g., in Table 3 where AAR for study participants is 1%, NNT is  $1 \div .01 = 100$ ).

A visual representation of NNT has been suggested as a way to make this metric even more concrete. As shown in Figure 2, of 100 patients who have otitis media:

- 86 would have been free from pain at days two to seven regardless of intervention, providing the basis for the baseline risk calculation Clinical Event Rate (green faces)
- Nine experienced pain but were not helped by an Rx (red faces)
- Five experienced pain and were helped by an Rx (yellow faces), so five out of 100, thus an NNT of 20.



**Figure 2.** Number Needed To Treat (NNT) is an innovative approach to express baseline event rate and absolute risk reduction.

Of course, NNT is not a panacea. There is no magic NNT number to indicate when a therapy is warranted. Nonetheless, NNT (or its safety complement NNH, Number Needed To Harm) is one way to express baseline event rate and absolute risk reduction in a vivid patient-centric context.

Another example of a unique and clinically salient metric is Individual Net Benefit. This metric attempts to address the disconnect that often occurs between clinical trial results and applied clinical practice. This disconnect occurs because patients who fulfill criteria for inclusion in a clinical study are often more homogeneous than those suffering from the condition in the broader population; thus, the benefit of a medication to any given individual is sometimes difficult to gauge. A model suggested by medical researchers Lubsen and Tijssen,<sup>12</sup> and subsequently echoed by others,<sup>13</sup> enables a prediction of benefit (or harm) to a specific individual by drawing on information from both clinical trial and epidemiology sources and takes the general form:

$$\text{Individual Net Benefit} = \text{Absolute Risk Level} \times \text{Relative Risk Reduction} - \text{Adverse Events}$$

The details of this approach are beyond the scope of this article, but we mention it as another illustration of a creative and integrative metric that can potentially help advance discussion of clinical trial risk assessment.

Statisticians continue to debate some of the technical aspects of NNT, Individual Net Benefit, and other metrics intended to express absolute risk levels, while nevertheless broadly agreeing on the importance of metrics like these that do address absolute risk levels in some form.<sup>14, 15, 16</sup>

**Reaching a tipping point?**

Actionable clinical trial data is vitally important to help support the inevitable qualitative judgment that goes into any risk/benefit assessment for a medication. It is ultimately in the best interest of all players in the health care system to promote and encourage efforts for better and clearer reporting of clinical trial data—an observation that may seem obvious but is nonetheless difficult to implement without some central force behind it. It appears possible that the Avandia situation may push things to a tipping point that will provoke the FDA to put some real energy and resources into this effort.

In September 2006, the Institute of Medicine issued a comprehensive report on drug safety that included recommendations to the FDA<sup>17</sup> for improved public communication with a specific emphasis on risk communication—recommendations with which FDA

agreed.<sup>18</sup> On the heels of the Avandia coverage, in June 2007 FDA announced the formation of a committee of outside specialists designed to advise the agency on how to inform the public about the risks and benefits of medications. Fifteen experts were named to the Risk Communication Advisory Committee in November 2007, with its first meeting slated for later this month.<sup>19</sup>

We certainly support and cheer such activities by FDA and others, and have some optimism that these efforts may bear fruit. In contrast to other recent high-profile industry reform suggestions (e.g., comprehensive publication of trial data and increased use of head-to-head trials), adoption of more applied and intuitive metrics for risk/benefit reporting of clinical trial results appears to be a relatively more manageable task. Further, application of enhanced reporting metrics would have

**We support activities by FDA and others to improve risk communication, and have some optimism they may bear fruit.**

clear benefits across the board for all players in the health care system in the form of improved patient care, more transparent medication assessments, higher quality public discourse, and enhanced industry reputation. Going beyond the strictly clinical realm, more effective communication about medication risk/reward might also help pave the way for more productive discussions on how to best balance clinical and economic concerns in our health care system—a topic on which the surface has just been scratched in the U.S. market.

The adoption of more intuitive reporting metrics for clinical trial data is not offered as a cure all for confusion over drug safety and efficacy. Indeed, it would be naive to think that such a complex problem can be resolved through any one action. This effort, however, in our opinion, represents one potentially important and broadly beneficial step in the improvement of public discourse about our medical products.

**References**

1. S.E. Nissen and K. Wolski, "Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes," *New England Journal of Medicine*, 356, 2457–2471 (2007).
2. B. Meier, "For Drug Makers, a Downside to Full Disclosure," *New York Times*. (23 May 2007).
3. Political Peer Review, *Wall Street Journal* A16 (18 June 2007).
4. R. Mathur and M.C. Stoppler, "Avandia—An Endocrinologists' Perspective," *MedicineNet.com* (accessed June 22, 2007).
5. R.L. Krall, "Cardiovascular Safety of Rosiglitazone," *The Lancet*, 369, (9578) 1995–1996 (2007).
6. K.A. Katz, "The (Relative) Risk of Using Odds Ratios," *Archives of Dermatology*, 142, 761–764 (2006).

7. J.C. Sinclair and M.B. Bracken, "Clinically Useful Measures of Effect in Binary Analyses of Randomized Trials," *Journal of Clinical Epidemiology*, 47, 881–889 (1994).
8. C.D. Naylor, E. Chen, B. Strauss, "Measured Enthusiasm: Does the Method of Reporting Trial Results Alter Perceptions of Therapeutic Effectiveness," *Annals of Internal Medicine*, 117, 916–921 (1992).
9. A. Johnson, "A Risk in Cholesterol Drugs is Detected, But is it Real?" *Wall Street Journal* A1, A14 (3 July 2007).
10. K. Stark, "Considering Alternatives," *Philadelphia Inquirer* C1, C7 (13 June 2007).
11. A. Laupacis, D.L. Sackett, R.S. Roberts, "An Assessment of Clinically Useful Measures of the Consequences of Treatment," *New England Journal of Medicine*, 318, 1728–1733 (1988).
12. J. Lubsen and J.G.P. Tijssen, "Large Trials with Simple Protocols: Indications and Contraindications," *Controlled Clinical Trials*, 10, 151–160 (1989).
13. P. Glasziou and L. Irwig, "An Evidence Based Approach to Individualizing Treatment," *British Medical Journal*, 311, 1356–1359 (1995).
14. J.J. Deeks, "Issues in the Selection of a Summary Statistic for Meta-analysis of Clinical Trials with Binary Outcomes," *Statistics in Medicine*, 21, 1575–1600 (2002).
15. J.L. Hutton, "Numbers Needed to Treat: Properties and Problems (with Comments)," *Journal of Royal Statistical Society, Series A*, 163, 403–419 (2000).
16. L. Smeeth, A. Haines, S. Ebrahim, "Numbers Needed to Treat Derived From Meta-Analyses—Sometimes Informative, Usually Misleading," *British Medical Journal*, 311, 1548–1551 (1999).
17. Committee on the Assessment of the US Drug Safety System, *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, A. Baciu, K. Stratton, S.P. Burke, eds. (National Academies Press, Washington, D.C. 2007).
18. Food and Drug Administration, *The Future of Drug Safety: Promoting and Protecting the Health of the Public: FDA's Response to the Institute of Medicine's 2006 Report*, <http://www.fda.gov/oc/reports/iom013007.html> (accessed June 30, 2007).
19. Food and Drug Administration, "FDA Announces New Advisory Committee to Address Risk Communication," <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01739.html> (accessed January 4, 2007).

**R. Brian Attig**, \* PhD, is vice president, quantitative services division, *Psyma International Inc.*, 661 Moore Road, Suite 120, King of Prussia, PA 19406, email: [brian.attig@psyma-usa.com](mailto:brian.attig@psyma-usa.com). **Alison Clabaugh**, MA, is project manager/consultant, *Psyma International Inc.*

\*To whom all correspondence should be addressed.

 **FIND OUT MORE**  
 For more articles on statistical issues in clinical trials visit [actmagazine.com](http://actmagazine.com)