

# Observations from the OMERACT Drug Safety Summit, May 2008

LEE S. SIMON, C. VIBEKE STRAND, MAARTEN BOERS, PETER M. BROOKS, DAVID HENRY, and PETER S. TUGWELL

**ABSTRACT.** Due to mounting concern about determination of benefit and risk in the context of product development and clinical practice the OMERACT Executive identified the need to bring together a variety of specialists to define risk. At the Drug Safety Summit held at OMERACT 9, specialists spoke on their given topics and the group considered risk in the context of formally posed questions. (J Rheumatol 2009;36:2110–3; doi:10.3899/jrheum.090590)

*Key Indexing Terms:*

CLINICAL TRIALS DRUG SAFETY RISK ASSESSMENT PHARMACOEPIDEMOLOGY

Worldwide, there is great concern about the determination of benefit and risk in the context of both product development and clinical practice. In view of this the OMERACT Executive identified the need to bring together clinical trialists, pharmacoepidemiologists, clinicians, clinical epidemiologists, statistical experts, and regulatory representatives to discuss different approaches to define risk. Each attendee spoke on a given topic and the group was charged to consider the issue of risk in the context of formally posed questions. Background context and an introduction to the Drug Safety Summit is provided. A companion article summarizes the presentations and discussions that followed<sup>1</sup>.

## BACKGROUND

There is a general perception that an approved therapeutic agent should not induce harm, and if harm is recognized, that the system developed to protect the consumer may appear to be flawed. When this concern is exacerbated by media attention and by other actions and reports, trust in the pharmaceutical and biotechnology industry and in the oversight regulatory organizations can erode, with perception of the behavior of sponsors and/or others ranging from negligent to reprehensible. The resulting “stew” can become quite toxic. Add political dimensions, and we find the cur-

rent situation of concerns and mistrust in the integrity of the regulatory process.

As a society we must ask ourselves whether our expectations and demands are fair or appropriate. In the United States and elsewhere, part of this problem has arisen because society has abrogated some of its responsibilities: research to help the public understand safety is underfunded; design and conduct of randomized controlled trials is relegated to sponsors, and there is no funding for a public effort to ascertain from a societal perspective relative benefit versus risk profiles in the context of similar products. The public’s need for information on comparative efficacy and safety of a class of therapeutics, for example, can never be ascertained by the efforts of individual sponsors of those products. Society has not provided necessary support to regulatory agencies charged with ascertaining relative benefit versus relative risk of new therapeutics (there are only varying levels of risk since there are no absolutely safe drugs in any one patient); moreover the monitoring of accumulating postmarketing evidence vis a vis consumers’ expectations is underfunded.

Sponsors are responsible to stockholders, and in a capitalist economy their interests may be incongruent with those of society in general. Everyone wants excellent health care and accurately assessed, safe and effective therapies, but sponsors should not, and cannot be the only source of data to inform decisions regarding safety. Ascertaining the information that consumers want and deserve should not be made the sole responsibility of sponsors, as this will not foster an accurate system for assessing benefits versus risks of new therapeutics. Regulatory authorities cannot be held accountable if society will not offer appropriate resources to establish appropriately detailed datasets to answer important risk-benefit questions.

## CONSIDERATION OF RISK

Risk is a critical consideration for everyone on a daily basis: The risk of driving to the store in bad weather versus the

---

*From the Harvard Medical School, Boston, Massachusetts; Division of Immunology, Stanford University, Palo Alto, California, USA; VU University Medical Center, Amsterdam, Netherlands; Faculty in Health Services, University of Queensland, Royal Brisbane Hospital, Herston, Australia; Institute for Clinical Evaluative Services, Toronto; and the Centre for Global Health, University of Ottawa, Ottawa, Canada.*

*L.S. Simon, MD, Associate Clinical Professor of Medicine, Harvard Medical School; C.V. Strand, MD, Clinical Professor of Medicine, Adj. Division of Immunology, Stanford University; M. Boers, MD, Professor of Clinical Epidemiology, VU University Medical Center; P.M. Brooks, MD, Faculty in Health Services, University of Queensland, Royal Brisbane Hospital; D. Henry, MD, President and CEO, Institute for Clinical Evaluative Services; P.S. Tugwell, MD, Director, Centre for Global Health, University of Ottawa,*

*Address correspondence to Dr. L.S. Simon MD, Associate Clinical Professor of Medicine, Harvard Medical School, Boston, Massachusetts, USA. E-mail: omeract@uottawa.ca*

benefit of shopping; balancing the risks of boarding a plane in bad versus good weather. Differentiating risk at an individual level compared with risk at a population level is difficult.

Multiple risks must be considered in the context of drug product development:

1. The initial risk to the sponsor of no return on investment if a therapeutic agent is not approved for marketing because it does not work as expected
2. A novel therapeutic intervention to improve signs and symptoms but not cure a disease carries an associated risk of an observed adverse event that would limit its use
3. Defining risk-benefit in a new therapy for use chronically that offers only symptomatic benefit and has associated adverse events resulting in irreversible damage may be easy, but if a therapy improved how a patient feels and functions, yet posed a risk of sustaining irreversible damage in 1/10,000 per year, would it be reasonable to use it? Would it even be approved? Further, if use of a new agent improves more than symptoms but does not provide a cure, does this change the above benefit/risk calculation<sup>1</sup>?
4. Risks to society include paying for the assessment of risks and benefits of a drug that does not work or causes more harm than good, or is no better than a much less expensive product, and paying for (part of) the associated health care.

The ability to assess benefit in the context of risk is difficult when measures of each component are so disparate. The risk incurred by exposure to a therapeutic agent as assessed in a clinical trial data set often can only be descriptive. Trials are usually powered to detect differences in efficacy, not toxicity, and in a highly selected population. In addition, minor imbalances remaining after randomization may have large (but unknowable) effects on the risk of one or more rare adverse events.

And if we design a trial to assess risk, how can effect sizes be estimated for a specific risk or group of risks? Estimates of effect sizes demonstrating efficacy are predicated on iterative data accumulated with increased patient exposure; thus, dose and dose duration can be appropriately defined. Similar estimates for risk are elusive.

The traditional “rule of threes” statistically defines safety or the upper limit of expected occurrence in the context of extent of exposure, in terms of overall numbers exposed, as well as duration of exposure<sup>2</sup>. Simply put, if it is desirable to understand a risk of an event at a rate of 1% then 300 patients need to be studied, while if there is interest to understand risk at a rate of 0.1% then 3000 patients need to be studied. The level of safety is expressed as the number of specified events per 100 or 1000 patient years of exposure using a cross-sectional database. However, to understand absolute level of effect it may be useful to define the incidence of major adverse events as an absolute number rather than relative risk. Thus there can be a better understanding

of the benefit to risk ratio, not risk relative to placebo or some other active therapy, but as a function of actual use of the agent.

Overall, the safety evidence required to allow acceptance of a new therapeutic agent is at least partially informed by the amount of benefit measured. If benefit, or efficacy, is mostly defined in the context of randomized controlled trials underpowered to accurately assess the incidence of rare adverse events, a partial solution is to perform large “pragmatic” trials.

Is the list of potential advantages and disadvantages of studying safety from clinical trial data complete? Do we want it to be complete?

Additional safety evidence can also be accumulated via postmarketing surveillance or observational data sets, although benefit is difficult to quantify in this context. The safety profile of a new agent can be further refined through signal accumulation via pharmacovigilance although assigning causality without consistent access to individual medical records is also difficult.

How to develop a system that allows all stakeholders to be best informed about the benefits and risks of a new therapeutic agent at any specific time in its life span? Are regulatory approvals and adequate product labeling sufficient to inform clinicians and/or patients that a new product is sufficiently safe in the context of its potential benefit? What role should each of the ways of measuring benefit and risk play in this process? Should efforts focus on developing new ways of measuring benefits versus risks? Or on better ways of communicating benefit/risk profiles of therapeutics so that interested parties can make better choices?

## ISSUES PRESENTED AS DISCUSSION ITEMS

1. Today the idea of relative risk versus relative benefit has been politicized and may not be well understood by many stakeholders. How might we improve this situation? We have successfully defined efficacy of treatments in several rheumatologic conditions but metrics to better define their safety are limited: (a) by required exposure to the new agent prior to approval as determined by International Conference on Harmonization (ICH) guidelines; (b) number of patient years of exposure required to define efficacy as well as safety; and (c) pragmatic considerations, such as overall population numbers, and currently available therapies. There are good metrics to define “number needed to treat” for benefit, based on data from randomized controlled trials but the “number needed to [potentially] harm” is not easily understood.

2. Because designing trials to confirm efficacy typically requires exclusion of patients with chronic comorbid conditions or low disease activity, those studied may not reflect the real world population that will be treated once a new agent is approved. This is predominantly true because trials are designed for regulatory approval, i.e., to demonstrate a

difference versus a comparator arm and with the fewest confounding issues possible to allow to determine efficacy.

In the past, risks of death or serious adverse events were based on the requirement of 2,500 to 3,000 patient years of exposure, necessary to identify rare adverse events. Recent efforts have focused on identification of any serious adverse event in terms of relative risks of 1 in 10,000, which requires exposure of 30,000 patients for chronic palliative noncurative therapies prior to approval. A more reasoned and pragmatic approach must be considered for pre-approval treatment exposure. Existing ICH guidelines may not be sufficient, but there should be a clear middle ground between the ICH requirements of a minimum of 1,500 versus 30,000 exposed. Even if large randomized controlled trials can be appropriately performed, it remains difficult to accurately determine sample sizes for defining the safety profile of a new therapeutic. Any specific therapeutic can only be considered safe in the context of the benefit it may offer. Further, it is difficult to calculate safety based on a mix of adverse events – not to mention determining causality in the context of comorbidities and concomitant therapies associated with the underlying disease. Should there be considerations given to employment of differing models or variations of the Naranjo algorithm<sup>3</sup>?

3. The ability of postmarketing surveillance data to inform these issues is limited, due to channeling bias, informed censoring, and changes in medical practice, and compounded by the difficulty to recruit into rigorous longterm trials once a new agent is approved, posing many difficulties for ascertaining safety signals.

#### QUESTIONS CONSIDERED AT THE SUMMIT

1. What are the limitations of a clinical development program to define rare risks and should this be addressed?

2. Can postmarketing surveillance be constructed to improve identification of signals regarding risk?

The following options might be applied to answer this question: proactive Risk Management/Pharmacovigilance planning processes, introduction of risk minimization tools, and evaluation of their impact into longterm extension trials, conducted also for claim of prevention of disability, particularly in rheumatoid arthritis, but perhaps in other clinical states that require longterm efficacy RCT analyses.

3. Once a safety signal is identified, are there better methods of assigning causality?

4. What is the utility of registries to define risk?

5. What is the utility of large simple trials to define risk?

Challenges exist regarding the construction of comparator groups.

How do we sufficiently estimate or determine an effect size to design an adequately powered trial to define safety and particularly when there are competing risks? How do we rule out the occurrence of a defined adverse effect in the context of many safety signals?

6. What is the utility of metaanalyses to define risk? Are they useful for identifying a “signal” or do they demonstrate proof of an adverse event?

7. Based on current concerns and the above issues: Should there be a rationale for conditional approvals for any new therapy with expectations that a registry be created for each new treatment to provide stringent, robust followup after approval with “real world” use information? How likely is it that these types of data sets will demonstrate improved information regarding safety of a new treatment versus what is presently available?

If the drug were conditionally approved with the requirements that it be studied more rigorously in the postapproval environment, then present RCT designs to prove efficacy and data sets to determine safety signals within the defined studied patient population could be augmented with a more “real world” approach. That would mean including a broader, more realistic patient population, more reflective of who might use the therapeutic, in terms of comorbidities, polypharmacy, age, and other characteristics already discussed. Such advantage is offset, however, by problems of adherence and recruitment into clinical trials, once a drug has been made available on the market.

8. Must we be dependent on pharmacoepidemiologic studies of larger populations?

In that context are patient years of exposure the right method to ascertain individual patient risk? If so, how do you determine causality, and if it is necessary to convert to patient years of exposure what is more important in signal generation, RCT data with embedded comparator rates or standard incident rates?

9. Should we develop a single metric that assesses both benefit and risk into a single metric (or rank)? If such a metric is developed what is its purpose? Would it be applicable for regulatory approval? Would be it used only for labelling purposes with no regulatory threshold?

10. Should other models from other disciplines be adapted to measure risk?

The discussions regarding the last several questions were considered to address the development of a research agenda. First and foremost was the realization that certain approaches would require an examination of the present methodology to interpret derived evidence. Overall the group determined that it would first approach registries to examine their present state of development and to begin to develop consensus requirements for both establishing a registry to both maximize the information to be initially derived and to develop the appropriate methodological approaches to analyze the data.

The conference ended at this point with the plan to pursue the issue of registries noted above. Furthermore, there was a clear commitment to continue the discussions regarding the other important questions.

## ACKNOWLEDGMENT

Drug Safety Summit Participants: Maarten Boers, Claire Bombardier, Peter Brooks, Maxime Dougados, Ralph Edwards, James Fries, Dan Furst, Larry Goldkind, Gord Guyatt, David Henry, Kent Johnson, Chris Kelman, Andreas Laupacis, Amye Leong, Larry Lynd, Tom MacDonald, Muhammad Mamdani, Andrew Moore, Pam Richards, Kenneth Saag, Jeff Siegel, Alan Silman, Lee Simon, Josef Smolen, Randall Stevens, Vibeke Strand, Anja Strangfeld, Miriam Sturkenboom, Samy Suissa, Peter Tugwell, Alan Tyndall, Vivian Welch, George Wells, Janet Woodcock, Thasia Woodworth.

## REFERENCES

1. Simon LS, Strand CV, Boers M, et al. Proceedings of the OMERACT 9 Drug Safety Summit: controversies and concerns about how to ascertain drug safety in the context of benefit. *J Rheumatol* 2009;36:2110-3.
2. US Food and Drug Administration Advisory Committee. Briefing information regarding Arava®; March 5, 2003. [Internet. Accessed May 14, 2009.] Available from: <http://www.fda.gov/ohrms/dockets/ac/03/briefing/3930b2.htm>
3. Javonovic BD, Levy PS. A look at the rule of three. *Am Statistician* 1997;51:137-9.
4. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981;30:239-45.
5. Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or Casual? The role of causality assessment in pharmacovigilance. *Drug Safety* 1997;17:374-89.
6. Atkins D, Best D, Briss P, et al, and the GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490-9.