Drug Safety 2008; 31 (11): 961-963
0114-5916/08/0011-0961/$48.00/0

Controlled Trials and Risk of Harm

I. Ralph Edwards

Uppsala Monitoring Centre, Uppsala, Sweden

I regularly read and hear declarations claiming that spontaneous reports are the poorest form of epidemiological evidence and, conversely, that controlled clinical trials provide the only evidence to base decisions on. Two statements in support of the latter position recently appeared in the *New England Journal of Medicine*, "The randomised clinical trial is considered to be the most reliable tool to assess the efficacy and safety of new drugs"[1] and the *British Medical Journal*, "Only properly randomised controlled trials can provide truly reliable evidence on adverse events, just as they are the only source of convincing data on drug efficacy."[2] Also, Aronson and Hauben[3] recently found it necessary to qualify the statement, "Anecdotal reports, by which we mean either individual cases or small case series, are generally regarded as providing poor quality evidence."

I continue to be surprised that, given the well known nature of the tools we use to look at the potential positive and negative effects of drug products, there are still fruitless controversies about what is overall 'best', rather than a much needed synthesis of viewpoints.[4]

Randomized controlled trials are not only by definition controlled, and tightly controlled, but they are also restricted in other ways. Observational, non-interventional studies make observations about treatments as they are, or have been, used in routine medical practice, which includes all the variables that are present in medical practice. The control group used for observational studies needs to mimic the treatment group including a range of variables, and taking into account all the variables present in real life is problematic.

Individual case safety reports (ICSR) are reports of concerns in single patients. Rather unhelpfully, ICSR are commonly called spontaneous reports, whereas more precisely they should be termed 'individual case harm reports'. They are the product of the experience and diagnostic logic of the reporter. ICSR are not epidemiology: even if the reports are aggregated, quantitative analysis only serves to find disproportionally reported excesses of drug/suspect adverse drug reaction combinations.[5-8] ICSR have no quantitative value other than to indicate that a number of reporters have the same concern, and ascertaining causality lies in the evaluation of the clinical decision-making logic behind them.[9]

So what should we deduce from the brief descriptions given here in how we view these tools?[10]

1. Efficacy and Hazard

If we want to know whether a drug has any specified effect in patients we should define the relevant variables in the patients as tightly as possible, and have a tightly matched control group. The treated group is artificial in so far as it is not representative of the population with the treated disease. The usual focus on only one defined variable for 'efficacy', that is, does the treatment work the way we predict it will, precludes definitive conclusions being drawn about other outcomes. Since both the duration of, and the number of patients enrolled in, each study are limited, this leads to a severe restriction when it comes to the ability of the study to determine harm and to predict risk or longer term effectiveness. Any harm that is seen during a clinical trial will be only the most common, and strictly, since the studies generally do not have harm as an intended hypothesis, a chance finding or confound-

ed association that must be seen as an alternative hypothesis. More worrying though, is that the power of most clinical trials is too limited to detect harm to a level that most would regard as relevant.

In summary, the value of the clinical trial is to determine that the drug may have an effect in humans (albeit often a surrogate marker effect), which can be defined as 'efficacy', and that there may be some indication of a possible common range of harm, which should be defined as 'hazard', since much more needs to be learned after the drug is marketed and the total risk profile becomes more evident.

2. Effectiveness and Risk

Large observational prospective studies are better equipped at showing the overall 'effectiveness' and 'known risk' profile, but even so are restricted in their ability to identify unexpected and rare events. Moreover, the more defined the exposed population, the easier it is to determine controls but the smaller and less representative of the patient population the study population is. The longer the study the more extraneous, and probably unpredicted, variables will creep in. Studies can be made as large as resources and time permit and, in principle, any number of outcomes from a single exposure can be investigated, but any finding that is not predefined should be regarded as tentative.

Case-control studies start with a defined hypothesis on a single proposed 'harm' and control for exposure variables of interest and exclude cases with known biases or confounding. Multiple exposures can also be evaluated. Finding suitable controls may be a difficult task.

It is clear that in clinical practice drugs are used for indications that do not fit tightly controlled definitions, patients have a range of severity of disease, medical errors in diagnosis or dosing may occur, patients may not take the medication as supposed, patients may take other drugs and have concomitant diseases, and so on. All of these complications may be exacerbated by data that are incomplete or which are simply recorded inaccurately: this may apply to both prospective and retrospective studies, but the world of epidemiology is developing better techniques to manage complex heterogeneous data all the time.

What one can get from observational studies is therefore a view of 'effectiveness' in real life use, as well as specified 'harms' and a quantification and confirmation of 'known risk' signals.

3. Benefit and Harm

ICSR have the widest coverage of exposed populations and therefore provide the best chance for alert professionals and consumers to voice their concerns, which can then be developed into hypotheses on 'unknown harms' for further study. Only data mining of healthcare records, which we are using successfully on an experimental basis, has anything like the same potential for hypothesis generation.[11] Sometimes individual cases can be extremely persuasive of drug-related harm, but even then quantification needs other work. What we should get from ICSR is qualitative information. We could also get better descriptions on how patients really feel 'benefit' from the drug, not only that it improves some pathology, physiology or symptom. In the same way, we should have a description of how the patient is 'harmed' and what that feels like. This information is invaluable to the development of meaningful quality-of-life measures, rather than professionals second guessing what benefit and harm feels like for a person through rating scales, free text descriptions and other indirect measures.

4. The Way Forward

Instead of trying to say which approach or which type of data is 'best', we should spend time trying to understand why results differ when different types of study are used. ISCR may give convincing information on rare cases of harm that formal studies do not have the power to detect. Observational data that shows less effectiveness and/or more harm than seen in clinical trials might indicate poor patient selection or a number of other clinical matters such as adherence issues that should be managed. In each of these situations the best we can do is to use our wisdom to

understand what might have happened, make reasonable changes and evaluate the outcome.

Life is really Bayesian[12] and Popperian[13] in the end: no study is definitive but changes prior knowledge giving a new testable hypothesis. As we get more information and understand the reasons for similarity and difference, so we get closer to a situation where reasoned action is possible, always taking into account other matters such as the seriousness of the situation, possible risk groups and available alternative disease management options. It is regulatory decision making throughout the process from signal to action that should be under review and discussion, not just the methods we use. We should also realize that once a decision is made, the impact needs assessment. More than anything, we need to be able to justify this whole process to the public, the media and in courts of law.

Acknowledgements

No sources of funding were used to assist in the preparation of this editorial. The author has no conflicts of interest that are directly relevant to the content of this editorial.

References

1. Drazen JM, D'Agostino RB, Ware JH, et al. Ezetimibe and cancer: an uncertain association. N Engl J Med 2008; 359 (13): 1398-9
2. Freemantle M, Irs A. Observational evidence for determining drug safety: is no substitute for randomised controlled trials. BMJ 2008; 336 (7645): 627-8
3. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. BMJ 2006; 333: 1267-9
4. Vandenbroucke J. Observational research, randomised trials, and two views of medical science. PLoS Med 2008; 5 (3): 339-43
5. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998; 54 (4): 315-21
6. Bate AJ, Lindquist M, Edwards IR, et al. Understanding quantitative signal detection methods in spontaneously reported data. Pharmacoepidemiol Drug Saf 2002; 11 Suppl. 1: S214-5
7. Bate A, Edwards IR. Data mining in spontaneous reports. Basic Clin Pharmacol Toxicol 2006; 98 (3): 324-30
8. Lindquist M. Use of triage strategies in the WHO signal-detection process. Drug Saf 2007; 30 (7): 635-7
9. Edwards IR. Spontaneous reporting – of what? Clinical concerns about drugs. Br J Clin Pharmacol 1999; 48 (2): 138-41
10. Edwards IR, Lindquist ML. Understanding and communication of key concepts in therapeutics. In: Moments of truth: communicating drug safety. Procedings of a Symposium in honour of Dr. M.N.G Dukes; 2000 Sep 22; Verona. Amsterdam: Elsevier, 2000: 9-14
11. Norén GN, Bate A, Hopstadius J, et al. Temporal pattern discovery for trends and transient effects: its application to patient records. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24-27; Las Vegas (NV)
12. Bayes T. An essay towards solving a problem in the doctrine of chances. Phil Trans R Soc Lond 1763; 53: 370-418
13. Popper K. Conjectures and refutations. London: Routledge and Kegan Paul, 1963

Correspondence: Professor *I. Ralph Edwards*, Uppsala Monitoring Centre, Stora Torget 3, Uppsala, S-75320, Sweden. E-mail: ralph.edwards@who-umc.org