Symmetrical analysis of risk–benefit

John B. Warren,¹ Simon Day² & Peter Feldschreiber³

¹Medicines Assessment, 196 Rotherhithe Street, London SE16 7RB, ²Roche Products Ltd, Welwyn Garden City AL7 1TW and ³Barrister, 4 New Square, Lincoln's Inn, London WC2A 3RJ, UK

Correspondence

Dr John Warren MD FRCP, Medicines Assessment, 196 Rotherhithe Street, London SE16 7RB, UK.
Tel.: +44 07789 825 680
E-mail: info@medicinesassessment.com

Keywords

efficacy, myocardial infarction, pharmacovigilance, regulation, safety

Received

16 November 2011

Accepted

28 February 2012

Accepted Article Published Online

8 March 2012

To quantify the value of a medical therapy the benefits are weighed against the risks. Effectiveness is defined by objective evidence from predefined endpoints. This benefit is offset against the disadvantage of adverse events. The safety assessment is usually a subjective summary of concerns that can often be neither confirmed nor dismissed. But sometimes a clinical database is so large that a parameter common to both efficacy and safety can be quantified with reasonable certainty: myocardial infarction (MI) is used here as an example. Recently the Food and Drug Administration (FDA) proposed set limits for the incidence of MI as a safety threshold for diabetes treatment. Setting a threshold before something is considered as a safety concern opens the possibility of setting a threshold for clinically important efficacy. When a parameter is common to both safety and efficacy, then logically a unit change in either direction should be of equal weight in the risk and benefit analysis. For example, a doubling in the incidence of myocardial infarction as a safety signal should be given equal weight to the halving of the incidence of myocardial infarction as an efficacy signal. Similarly, if FDA guidance suggests that a less than a 30% increase in the incidence of MI as a safety parameter is considered acceptable, for example for diabetes treatment, when there is no other major toxicity, this opens a debate about a possible inverse threshold for clinical benefit for drugs that reduce a risk factor, such as antihypertensives.

Introduction

Much has been written about the objective evaluation of the risks and benefits of medicines. Whilst such analyses should be fair when weighing evidence in favour of and evidence against the use of a drug, equitable judgement is confounded when the methodology of the two aspects of this assessment differ. Different methodologies for safety and efficacy assessments are usual when evaluating new medicines. The wide range of risks associated with, for example, warfarin or anti-cancer medication, are measured in quite different ways to the quantification of their benefits. There is no single scale of a continuous parameter, other than mortality, that summarizes risk–benefit and considering the symmetry of one parameter would be inappropriate. Disparate events, such as immunosuppression and gastrointestinal toxicity, might be weighed against tumour progression free survival. Risk–benefit analyses are rarely as simple as weighing one extra myocardial infarction (MI) as a risk against one less MI as a benefit.

At the time of licensing the key proof of efficacy comes from pre-defined trial endpoints that are ranked in order of importance. This pre-definition is needed to control the type I error, to reduce the risk of claiming an effect that is not real. In contrast, most safety assessments involve a retrospective trawl of multiple potential signals. The multiplicity of safety signals usually excludes the possibility of pre-defining trial safety end points by rank, making valid statistical significance testing difficult, if not impossible. These analyses, of course, depend on the quality of the trial data and to detect a change in the incidence of MI requires controlled clinical trials, given that MI is often a common background event in elderly or diabetic trial populations.

Recently there has been a shift in regulatory thinking stimulated by the Vioxx safety issue of 2004 [1]. It became useful to quantify the comparative risks of different COX2 inhibitors at different doses in terms of cardiovascular risk. This raised the possibility of setting a limit for one aspect of safety, the odds ratio for cardiovascular disease, above which a drug should be withdrawn. Such a ratio only contributes to part of the decision as COX2 inhibitors have

anti-inflammatory efficacy with a reduced incidence of gastrointestinal bleeding, advantages that may outweigh some of the cardiovascular risk.

The cardiovascular odds ratio regulatory debate intensified when rosiglitazone was found to be associated with cardiovascular adverse events [1]. Treatment of type 2 diabetes aims to reduce the complications of hyperglycaemia and to lower the incidence of cardiovascular complications. Instead the rosiglitazone data indicated a potential increased cardiovascular risk. This is a simpler issue than the risk–benefit of COX2 inhibitors, as glitazones have few additional benefit or safety issues to complicate the key effect on cardiovascular events when calculating risk–benefit. The expected benefit of glitazones is largely confined to a lowering of blood glucose that should translate into a reduction in the incidence of the cardiovascular complications of type 2 diabetes. For the glitazones, regulators were obliged to discuss a threshold risk of MI that might be acceptable. Though the EU has not fixed such acceptance levels for increased MI risk, the FDA guidance has done so for this class of compound [2]. This FDA advice sets a lower threshold for risk than that set in many previous US court cases, where up to a doubling of risk has been found acceptable. The reason for this probably relates to the reliability of the data. A greater margin of error is to be expected for observational data and a threshold of two was used during the debate on the risks of oral contraceptives where most of the evidence came from epidemiology and confounding factors led to uncertainty [3, 4]. We can be more certain of the incidence of MI from the data of large controlled trials.

When an end point can be pre-defined and made primary for both safety and efficacy, for example MI, particularly in a trial that is adequately powered, then the magnitude of change, either positive or negative, that is considered important should be evaluated consistently. This is an unusual example of risk–benefit assessment where the symmetry of the evaluation can be examined on a continuous scale. Another example of such a scale is mortality, where much of the risk–benefit of a drug can be summarized as to whether it shortens, lengthens or has no effect on life expectancy. But for most medicines data on mortality are scarce and for this reason MI was chosen as the illustration for this debate.

Many ways can be used to quantify change with treatment, whether of benefit, or of an increase in adverse events. Such change is often expressed as an odds ratio, both for events that are prevented and for those that are increased. When considering causation, it is usual to refer to the nine conditions specified in Bradford Hill's criteria for causation [5]. The first of these is the strength of the association, which can be measured in terms of relative risk, or the odds ratio. The remainder address additional evidence that support the association.

An odds ratio has been recommended by the FDA to propose a numerical threshold for safety for anti-diabetic

therapy. If this approach is valid, then it may be logical to adopt a similar threshold for evidence of efficacy. If a medicine improves a biomarker such as glucose, with little other efficacy or safety considerations, then it may be reasonable to take the FDA approach and define a threshold below which any increase in cardiovascular disease is not considered important. But this opens the possibility of inverting this argument. If a medicine improves a biomarker such as glucose or blood pressure, with little other efficacy or safety considerations, then it may be reasonable to take the inverse of the FDA approach and define a threshold below which any decrease in cardiovascular disease is not considered important. The odds ratio has the advantage that it is symmetrical about 1, so that an odds ratio of harm of 2 is mirrored by the odds of benefit of $1/2$.

Quantifying efficacy

Many drugs have been approved with small, but statistically significant, improvements in efficacy. Though the evidence of efficacy may be compelling, the effect size in terms of the odds ratio can be small. Using MI as an example, it seems likely that the odds ratio for MI reduction by amlodipine is about 0.93 compared with control (95% CI 0.87, 0.99) [6]. Similarly the reduction in MI with angiotensin receptor blockers (ARBs) compared with control gives an odds ratio of 0.99 (95% CI 0.92, 1.07) [7]. Of course there are additional benefits to ARBs other than a lack of effect on MI. They reduce the odds ratio for stroke to about 0.9, for heart failure to 0.87 and new onset diabetes to 0.85. Although blood pressure is reduced, the benefits of amlodipine and ARBs are small, given that the aim of hypertension treatment is to prevent cardiovascular disease and they would make little impact on the 1100% increased risk of MI caused by smoking 40 cigarettes per day [8] or even the 50% increase in risk of MI with passive smoking [9]. Hypertensive patients are about 4–5 times more likely to have a MI than a stroke, though stroke may be the more disabling. Calcium channel blockers and the ARB group of drugs are included with angiotensin converting enzyme inhibitors in UK guidelines as drugs of first choice for hypertensive patients younger than 55 years of age [10], yet the cardiovascular benefits are considerably less than a reduction of 20%, that is the odds ratios for benefit are >0.8 .

Quantifying safety

A patient who takes amlodipine, or an ARB, to reduce their chance of an MI might also take a drug with an effect on MI that is more than likely to negate such benefit. For example, it has been proposed that the odds ratio for MI with celecoxib is greater than 2.0 [11, 12], a risk which

might not be that different from the poorly defined cardiovascular risk of paracetamol (acetaminophen) [13].

Concerns about cardiovascular risk with the cyclooxygenase group of drugs recurred with reviews of cardiovascular events associated with glitazone treatment for diabetes. The FDA opinion on cardiovascular risk with anti-diabetic therapy [2] was triggered by adverse events associated with rosiglitazone use. This opinion sets an upper bound of a 95% CI for the estimated odds ratio for cardiovascular events of 1.8 premarketing and 1.3 post-marketing. This represents an important change in regulatory thinking, that an odds ratio below a set threshold, in this case for cardiovascular events, may be necessary as important evidence of lack of harm. The circumstance of this advice is unusual in that the expectation is for cardiovascular events to be decreased, not increased, by therapy and is not applicable to more complex risk–benefit evaluations.

It is rare for the safety evaluation to pre-define fully analyses in trial protocols. Many safety concerns are identified early in development and often halt a drug's progress. Progression to phase 3 implies that safety is considered acceptable to proceed to a wider population exposure and for this reason safety end points are not generally identified or ranked in advance in terms of their importance. The inability to pre-specify safety end points means that statistical multiplicity, and hence the type 1 error, cannot be adequately accounted for.

The need to identify important risk is balanced by the need not to waste resource on unnecessary concerns. Whilst regulators are expected to take a precautionary approach, this must be balanced and not unnecessarily cautious. As an example, small increases in the incidence of rash, liver and renal toxicity that do not reach statistical significance may be worth tolerating for a drug that significantly reduces an important primary efficacy end point, such as the incidence of pneumonia.

The safety assessment of new medicines usually refers to guidance on the long term treatment of non-life threatening conditions. The standard approach is defined by the International Conference on Harmonization guideline, ICH E1 [14]. This recommends about 1500 people should be exposed to a drug at the time of licensing, of whom some 300–600 are exposed for 6 months and at least 100 are exposed for 1 year. But when the signal might be an increase in cardiovascular events, much larger databases are required.

The odds ratio and numerical symmetry

Though attempts have been made to quantify the risk–benefit ratio objectively [15], it is usually a subjective judgement given the complexity of the safety assessment.

The situation changes when large clinical trial databases become available, usually post-marketing. Even if a cardiovascular biomarker is targeted, the underlying long term aim of treatment is to increase life expectancy, not to shorten it. If a drug is given to lower glycaemia in type 2 diabetes, then the expectation is that the incidence of MI will be reduced, not increased, as a consequence of better glucose control. In these cases the end points of mortality and myocardial infarction are common to the assessment of both the benefit and the risk.

Asking students to quantify the difference between percentage rates is a useful introduction to the risk–benefit of medicines. Take the example of two incidence rates of 1% and 2%. A difference between the two of +1%, or –1%, is symmetrical either side of no difference. The difference between the two can also be expressed as a 100% increase or a 50% decrease, margins which no longer have linear symmetry. Many find percentage differences uncomfortable with not everyone recognizing that a two-fold increase, i.e. a relative risk of 2.0, is synonymous with a 100% increase.

Suppose the chance of a MI is 1% in situation A and 2% in situation B. If A is when a cardiovascular drug is taken and B is with placebo, then this 50% reduction would signify benefit of greater magnitude than that achieved by any of the standard cardiovascular treatments for hypertension, lipid disorder or secondary heart disease prevention. The majority of cardiovascular drugs which are licensed to reduce the chance of a MI have smaller effects than this, suggesting that a proven halving of the incidence of MI would be widely interpreted as proof of outstanding efficacy.

But if A is with placebo and B is with a drug, would this then be interpreted as outstanding evidence of a safety concern? Human nature may accept good news more readily than bad. The seminal paper linking lung cancer with smoking in 1950 concluded 'The risk of developing the disease increases in proportion to the amount smoked. It may be 50 times as great among those who smoke 25 or more cigarettes a day as among non-smokers' [16]. This increase in risk is monumental when compared with any reduction in risk from therapeutic intervention, with the exception perhaps of vaccination. Yet the risk of smoking took many years to be recognized as casual by national and international authorities.

For symmetrical risk–benefit analysis an odds ratio of 0.5 for benefit should be given equal weight to an odds ratio of 2.0 for risk, though this statement needs qualification. If the background number of adverse events is 100 cases, an odds ratio of 0.5 will reduce the number of cases by 50 and an odds ratio of only 1.5 will increase the number of cases by 50. But if two drugs are taken together and one halves background incidence with an odds ratio of 0.5, this would be neutralized by a drug with an odds ratio of 2.0, but not by one with a ratio of 1.5. Whereas most drugs receive regulatory approval for benefit when the

relative risk reduction lies between 0.5 and 1.0, safety analyses in past US litigation has often limited causation to those cases where the relative risk point estimate, rather than a confidence interval, is greater than 2.0 [17]. This discrepancy arises partially from the greater uncertainty associated with observational safety data, compared with the more certain clinical trial data used for efficacy evaluation. But this observational data uncertainty is reduced when the incidence of MI, for example, with COX2 inhibitors or glitazones, is ascertained from clinical trial evidence.

Proof of causation in the courts

The effect size, as part of Bradford Hill's criteria of causation, is of considerable interest in court cases where potential harm from medicines is a matter of litigation. Many court cases have been determined by accepting that 'The threshold for concluding that an agent was more likely than not the cause of an individual's disease is a relative risk greater than 2.0' [18]. The drawbacks of such a threshold for tort cases have been debated [19, 20], with an increasing appreciation of the limitations of this approach.

That the drug is more likely to be guilty (odds ratio >2) than innocent (odds ratio <2) has an appeal to juries and has been used to make judgements quantifying the balance of probabilities in civil tort cases in some US trials [16]. In the UK this approach is under intense scrutiny and risks less than 2 are now being considered causal in some cases with supporting evidence.

Though a variety of methods can test the statistical probability of an association, it has been long accepted in epidemiology that in observational studies a RR of <2.0 is weak evidence of causation [15]. This is a sensible approach in so far as it limits chasing weak signals, but perhaps needs to be revised for controlled clinical trial evidence. An important caveat is that association, however large the relative risk, does not prove causation.

Conclusion

Whereas best guess point estimates guide efficacy judgements, expressed as the mean or odds ratio, safety judgements require the adoption of the precautionary principle and hence more consideration of an upper bound of a confidence interval.

The FDA requirement for evidence of reassurance on the cardiovascular safety of new anti-diabetic drugs [2] has important implications, namely that a pre-defined upper acceptance threshold has been proposed.

The FDA guidance has implications for efficacy and risk–benefit. If a threshold approach is taken for safety, it might be considered appropriate to adopt a symmetrical approach to efficacy data when the primary risk–benefit

end point is one continuous variable measured in large clinical trials. This has substantial implications for current cardiovascular drugs.

Competing Interests

JW worked previously as a medical assessor for the MHRA and helped represent the UK at the European Medicines Agency. He currently works as a consultant and has advised multiple pharmaceutical companies on drug development and regulation. None of the content of this article reflects a past or future conflict in terms of potential financial gain.

SD works in the pharmaceutical industry. He is a paid employee of Roche Products Ltd. He also has various stock options and pension plans invested with pharmaceutical companies. He worked previously for the MHRA as a statistical assessor, Head of Statistics and helped represent the UK at the European Medicines Agency. None of the content of this article reflects a past or future conflict in terms of potential financial gain.

PF is a medical assessor at MHRA. He is also a barrister at Four New Square, Lincolns' Inn. He has not and will not in the future make any financial gain from expressing the views in this publication.

REFERENCES

- Warren JB. The regulatory process: pharmacovigilance in practice. In: *The Law and Regulation of Medicines*, ed. Feldschreiber P. Oxford: Oxford University Press, 2008; 129–63.
- Food and Drug Administration. Guidance for Industry. Diabetes Mellitus – Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. 2008. 1–8. Available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf> (last accessed March 2012).
- McPherson K. Third generation oral contraception and venous thromboembolism. *BMJ* 1996; 312: 68–9.
- O'Brien PA. The third generation oral contraceptive controversy. The evidence shows they are less safe than second generation pills. *BMJ* 1999; 319: 795–6.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58: 295–300.
- Wang J, Li Y, Franklin S, Safar M. Prevention of stroke and myocardial infarction by amlodipine and angiotensin receptor blockers. A quantitative overview. *Hypertension* 2007; 50: 181–8.
- Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. *BMJ* 2011; 342: d2234.

- 8** Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004; 364: 937–52.
- 9** Teo KK, Ounpuu S, Hawken S, Pandev MR, Valentin V, Hunt D, Diaz R, Rashed W, Freeman R, Jiang L, Zhang X, Yusuf S, INTERHEART Study Investigators. Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. *Lancet* 2006; 368: 647–58.
- 10** National Institute for Health and Clinical Excellence. Hypertension. CG127, 2011. Available at <http://www.nice.org.uk/CG034> (last accessed March 2012).
- 11** Caldwell B, Aldington S, Weatherall M, Shirtcliffe P, Beasley R. Risk of cardiovascular events and celecoxib: a systematic review and meta-analysis. *J R Soc Med* 2006; 99: 132–40.
- 12** Food and Drug Administration. Highlights of prescribing information, Celecoxib. Available at http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/020998s033,021156s003lbl.pdf (last accessed March 2012).
- 13** Ritter JM, Harding I, Warren JB. Precaution, cyclooxygenase inhibition, and cardiovascular risk. *Trends Pharmacol Sci* 2009; 30: 503–8.
- 14** International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. The extent of population exposure to assess clinical safety for drugs intended for long term treatment of non-life threatening conditions. E1. Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf (last accessed March 2012).
- 15** Mussen F, Salek S, Walker SR. Benefit-Risk Appraisal of Medicines. A Systematic Approach to Decision-Making. Chichester: Wiley-Blackwell, 2009.
- 16** Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *BMJ* 1950; 2: 739–48.
- 17** Bailey LA, Gordis L, Green M, Rothstein PR II. Reference Guide on Epidemiology, in Reference Manual on Scientific Evidence. Washington, DC: Federal Judicial Center, 1994; 168. Available at <http://ftp.resource.org/courts.gov/fjc/sciam.6.epide.pdf> (last accessed March 2012).
- 18** Carruth RS, Goldstein BD. Relative risk greater than two in proof of causation in toxic tort litigation. *Jurimetrics* 2001; 41: 195–209.
- 19** Green MD. The future of proportional liability: the lessons of toxic substances causation. In: Exploring Tort Law, ed. Madden MS. New York: Cambridge University Press, 2005; 352–402.
- 20** Green MD, Freedman DM, Gordis L. Reference guide on epidemiology. In: Reference Manual on Scientific Evidence, 2nd edn. ed. Smith FM. Washington, DC: Federal Judicial Center, 2000; 333–400. Available at [http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/\\$file/sciman00.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf) (last accessed March 2012).