EDITORIAL

Drug Saf 2010; 33 (10): 805-809
0114-5916/10/0010-0805/$49.95/0

Reflections on Attribution and Decisions in Pharmacovigilance

Ola Caster and *I. Ralph Edwards*

Uppsala Monitoring Centre, Uppsala, Sweden

1. Causal Relationships

In general, a causal relationship depends upon the nature and amount of evidence supporting an attribution hypothesis, such that 'A causes B'. A can be a sufficient cause of B, meaning that A is always followed by B; or a necessary cause, meaning that B cannot occur without being preceded by A; or both. These deterministic concepts are not relevant to pharmacovigilance. No drug is a sufficient cause of an adverse effect, and there are no examples of necessary causes either. The trigger to modern pharmacovigilance, thalidomide, has come close to being considered a necessary cause of phocomelia, but this is not true. Phocomelia is very rare in the absence of thalidomide but not vanishingly so; x-irradiation is one example of another cause.[1] This example is illuminating also because it shows that the definition of the hypothesis is critical: one must be sure one knows what A and B are. In this case we have used the terms phocomelia, ectromelia, amelia and limb-reduction disorder loosely as synonyms when in reality they are overlapping hierarchical entities.

Lacking deterministic causality, we must rely on probabilistic evidence for our attribution hypotheses. Such evidence is present when the probability of B given A is 'large', and the probability of B given not-A is 'small'. Whereas this holds true for thalidomide and phocomelia, where the likelihood of a causal relationship is indeed very high, not even then is causality *certain*, and we must be aware of that. This requires those of us in pharmacovigilance to always consider the probability of one causality hypothesis against others. The real challenge is to prove a hypothesis wrong, which does not automatically succeed if there is support for a competing hypothesis because there can be multiple causes to an effect. Confounding is one very common example: if a competing hypothesis regarding the potential confounder C and its relation to B is accepted, for example by showing a statistically strong relationship between C and B, this does not automatically exclude the possibility of a causal link between A and B, though it could be that the probability of the causal link from C to B is greater. It may also be that A and C together have a higher likelihood of causing B via a synergistic or an additive effect, possibly even when A seems to have no effect on B in the absence of C.

2. Sources of Evidence

Pharmacovigilance in action is essentially Bayesian: a tentative prior probability for a hypothesis becomes modified up or down as further evidence is obtained, to become a posterior probability for new consideration. Such evidence may come from different sources and also the overall view of the data may use different logic. Individual case reports can provide excellent evidence on attribution in one or more specific patients. For example, positive dechallenge and rechallenge reactions, and the absence of other suspected agents (drugs or other), should result in a high *notional* probability for an attribution hypothesis that concerns only those particular

instances of drug intake. We argue that such evidence should guide also the probability for the *general* attribution hypothesis, which is most often our primary concern. Others have even suggested that there are examples of adverse effects for which well documented case series provide sufficient evidence to obviate further investigation.[2] Randomized clinical trials and observational studies follow a different logic. They can either refute the null hypothesis – meaning that the potentially causal relationship is unlikely to be due to chance – or else fail to do so. If the null is rejected, no matter how low the p-value might be, this does not establish causality in a deterministic sense. Furthermore, because of suboptimal choices of design or analysis methodology, such studies do not in general provide probabilistic evidence that the investigated attribution hypothesis is more likely than other hypotheses. Similarly, if the null is not rejected, this is insufficient to establish that there is not a causal relationship. The study might be underpowered or lack sufficient follow-up time. Therefore, we argue, a formal study that fails to reject the null cannot invalidate a finding in a strongly suggestive case series; it can only lower the posterior probability of the attribution hypothesis. The case series might also relate to representatives of some subpopulation, in which case there would have been a more appropriate overall attribution hypothesis for the formal study to investigate.

No matter what the source might be, the quality of the data is one important aspect, and certainly the Bradford-Hill criteria[3] must be part of an overview. These are summarized below, with our comments insofar as they are relevant to pharmacovigilance.

1. *Strength:* A weak association does not mean that there is not causality.
2. *Consistency:* Consistent findings observed by different persons in different places, with different samples, strengthen the likelihood of causality.
3. *Specificity:* Causality is more likely if the effect is observed in a very specific population at a specific geographic location and the disease has no other likely explanation.
4. *Temporality:* The effect has to occur after the cause and, if there is an expected delay between the cause and the effect, the effect must occur after that delay.
5. *Biological gradient:* A positive dose-response relationship strengthens the likelihood of a causal effect. With some interactions a negative dose-response relationship may be suggestive.
6. *Plausibility:* A plausible mechanism between cause and effect is an indicator of causality, but not all drug-effect mechanisms are known.
7. *Coherence*: Evidence from clinical laboratory or clinical pathology increases the likelihood of causality, but the same issue applies as in point 6: such evidence may be unavailable.
8. *Experiment:* Other experimental evidence such as animal studies may be supportive.
9. *Analogy:* The effect of similar factors may be important, such as class effects of drugs.

3. Tentative and Strong Signals

Evidence may be composed of heterogeneous information of different evidential weight and gathered over time. In pharmacovigilance, the first level of attribution hypothesis is generally referred to as a 'signal'. Neither available definitions of a signal[4] nor the description of a signal presented in a previous editorial in *Drug Safety*[5] provided any pharmacovigilance practice information. In this editorial we propose some decision-making thoughts for debate.

The defining property of a signal that we make use of here is that its evidence be sufficient to warrant further action of some sort. It seems reasonable that the likelihood of a signal representing a causal relationship should be one factor that influences the type of action. Following this logic, we distinguish between tentative (early) and strong signals. These are hypotheses whose *notional* probabilities of being true are below or above 50%, respectively.

As a general guidance as to what is a tentative signal and what is a strong signal, a small case series with high quality and informational content *might* reach the level of 50% notional probability, in spite of lacking the very important aspect of quantification, as well as enough detail that might explain the idiosyncrasies of the patients affected. However, such a signal should in most circumstances be regarded as tentative

because of these shortcomings.[6,7] Moreover, on the basis that many formal studies are concerned with excluding chance findings rather than comparing alternative hypotheses, as discussed above, a rejected null hypothesis in a single study should not generally *alone* result in a 50% or higher notional probability of a causal relationship. Often it will be the accumulation of individual case reports, weighted by their quality and content, in conjunction with formal studies and other background information – all cautiously considered in the context of the Bradford-Hill criteria – that will result in a notional probability above 50%. The coarse line of demarcation at 50% probability is introduced in view of our flawed and heterogeneous sources of evidence. For simplicity, in the argumentation to follow, this notional probability is treated as a point entity when in practice it would be valuable to account for its associated uncertainty interval, whose width will narrow down as more evidence accumulates.

4. Deciding on an Action

Under the assumption that we can classify a signal as either tentative or strong – as either more or less likely than not of being true – how does this affect our decision on best action, in particular with respect to issuing a warning immediately or pursuing further, active investigation? It would be reasonable to consider also the consequences that each alternative would have if the signal were true or false, respectively. In particular, it is relevant to consider the relative consequences of a type I error, to immediately report a false signal, and of a type II error, to delay reporting of a true signal in anticipation of firmer evidence to come. The former error will cause an unnecessary scare, whose consequences on patients' health will stem mainly from a decreased willingness to use the drug in question. The impact of this will depend on the type and magnitude of the benefit from the drug and whether that can be replaced by alternative therapies. It also depends on the degree to which such behaviour can be avoided through proper communication of the signal. The consequences of a type II error will depend, in a multiplicative manner, both on the seriousness or severity and duration of the adverse effect and the numbers likely to be affected. Also, the cumulative damage could be expected to increase with the length of delay in action. The heterogeneity of pharmacovigilance information does not allow for easy statistical probability calculations, and value judgements must be made on the contribution of each piece of evidence. Each such value judgement must be made transparently, and also with an audit trail to allow for precedents to be considered, leading to coherence and controlled flexibility in future situations.

If the situation allows for at least an ordinal assessment of the relative consequences of a type I and a type II error, this could be combined with the notional probability that the signal represents a causal relationship, to yield guidance on action. If probabilities and consequences are considered jointly, the following is implied: if the signal is tentative – that is, if we are more likely to be wrong than right – and the consequences of a type I error are worse than those of a type II error, then the signal should not be broadcast, but further investigated. Similarly, if the signal is strong and the consequences of a type II error are worse than those of a type I error, then immediate notification of the signal is recommended. Assessment is complicated by the fact that a decision to investigate the signal further will induce another similar decision in the future, whose evaluation in turn will depend on the outcome of the investigation itself. Furthermore, because considered and precise communication will reduce the negative consequences of a type I error, notional probabilities and consequences of actions will not, in general, be independent. For example, the weaker level of evidence attached to a tentative signal will often make such preventive communication more difficult, simply because we know less. Although there are many factors that matter, this should mean that type I errors are generally worse for tentative signals, and these signals should therefore be considered very carefully before warning. When case series of rare and specific clinical events are judged, for whatever reason, to be strong signals, the quality and content of information, and the type of event, should

enable effective communication and may make publication the preferred action.

Clearly it could happen that a type II error is considered worst for a tentative signal, or a type I error for a strong signal, in which case a more deliberate joint evaluation of strength of evidence and consequences of available actions may be necessary. The same holds true if immediate warning has been ruled out and there is more than one type of active investigation to consider. In general, a more thorough investigation will delay potential warning more, but will, on the other hand, provide firmer evidence that will hopefully settle the signal as either strong or as a non-signal. For example, tentative signals can be investigated by well constructed observational studies, though it may be reasonable in some cases, e.g. if the event is serious, to discount other competing hypotheses on the basis of an assessment of historical information and the Bradford-Hill criteria.

To complicate the picture further, some argue that decisions in pharmacovigilance should also be modified according to what regulatory action is feasible: one is less likely to decide to act early and actively if there is no subsequent preventative or mitigating action possible.[8] Often this argument is presented in the converse: act earlier if one can do something about the situation. This is seen usually in the public health perspective. Can one define at-risk groups? Is there any treatment? Can one define safer doses? The public health perspective tends to view idiosyncratic and/or rare adverse drug reactions (ADRs) as out of scope. This is a much too limited view of pharmacovigilance and makes no recognition of the importance of good ADR information in deciding which treatment to use *in an individual patient*, how a patient adheres to treatment if there are troublesome but non-serious ADRs, as well as how to consider ADRs in the differential diagnosis of a patient taking drugs. After all, most serious ADRs are rare illnesses in the orphan disease frequency range, the latter often defined as less than 1 in 1500, for each drug. However, the total burden of ADRs across many millions of drug exposures, each of which has the small risk, makes this a massive health problem.

5. Impact on Communication and Pharmacovigilance Conduct

Type I error is most likely to have the greatest effect in situations where a population is unprepared and where expectations of safety are high. Very well thought out communication is essential in pharmacovigilance to help mitigate the consequences from type I errors. Such communication will need to address the quantitative nature of the risk, and acknowledge the possibility of a misleadingly low population risk, if deduced from very preliminary data such as a few case reports. As we have tried to argue, evidence for one attribution hypothesis often competes with alternative hypotheses, such as the relationship might have occurred by chance, or that other diseases or drugs might have caused the event. In a given situation, there may be any number of 'plausible' as well as 'possible' hypotheses, with notional probability below or above 50%, respectively. Hence, there needs to be consideration on the identification of alternative hypotheses and an exploration of the relative strength of evidence between them. The competing hypothesis(es) with the strongest relative probability between several is most probable and should be communicated as such – provided a notification is the most appropriate action – though the other hypotheses must be mentioned as less likely. This approach follows the proposals of the Erice Declaration[9] and it ensures that there is an open presentation of all the probable causes and their likelihood with evidence and judgements on relative probabilities mentioned. Many epidemiological studies are designed and powered only to show that an evidential link to a single hypothesis has <5% statistical probability of being due to chance. Some studies investigate several competing hypotheses, and it is this comparison that has most importance for health professionals and patients in choosing treatments. Excellent communication will empower those who need to use the information, patients and health professionals, to make decisions and take actions that are important to them.

Acknowledgements

The authors have no conflicts of interest to declare that are directly relevant to the content of this editorial.

References

1. Galloway JL, Delgado I, Ros MA, et al. A reevaluation of X-irradiation-induced phocomelia and proximodistal limb patterning. Nature 2009; 460 (7253): 400-4
2. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. BMJ 2006 Dec; 333 (7581): 1267-9
3. Hill AB. The environment and disease: association or casuation? Proc R Soc Med 1965; 58 (5): 295-300
4. Hauben M, Aronson JK. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf 2009; 32 (2): 99-110
5. Edwards IR, Lindquist M. First, catch your signal! Drug Saf 2010; 33 (4): 257-60
6. Edwards IR, Lindquist M, Wiholm BE, et al. Quality criteria for early signals of possible adverse drug reactions. Lancet 1990 Jul; 336 (8708): 156-8
7. Kelly WN, Arellano FM, Barnes J, et al. Guidelines for submitting adverse event reports for publication. Drug Saf 2007; 30 (5): 367-73
8. Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. Pharmacoepidemiol Drug Saf 2003 Jan-Feb; 12 (1): 17-29
9. Bowdler J, editor. Effective communications in pharmacovigilance: the Erice report. Uppsala: The Uppsala Monitoring Centre, 1998

Correspondence: *Ola Caster*, Uppsala Monitoring Centre, Box 1051, S-75140 Uppsala, Sweden.
E-mail: ola.caster@who-umc.org