

7

Causality and Correlation in Pharmacovigilance

S. A. W. Shakir

Introduction and historical background

Human beings have been concerned about causality, i.e. what causes what in this world and why, for a long time. With regard to health, for thousands of years, men and women attributed events, such as disease, to a variety of physical and conceptual 'causes', including natural phenomena, e.g. an eclipse, evil spirits or the wrath of the Gods.

The publication in 1638 of Galileo's book *Discorsi* marks the beginning of the modern scientific era. In his book, Galileo introduced concepts such as description first, explanation second and that description could be carried out using the language of mathematics. Following Galileo, the most notable philosopher who wrote about causation was David Hume, the Scottish 18th century philosopher. In his major work *A Treatise of Human Nature*, Hume addressed a number of aspects related to causality, and many of the concepts that he proposed remain valid. One of his useful thoughts was that there is no such thing as an impression of a causal relationship. According to Hume, we can perceive by mere observation of *a* and *b* that *a* is above *b* or to the right of *b*. He held the view that when we say *a* causes *b*, we mean that *a* and *b* are constantly conjoined by fact, but not that there is some necessary connection between them. In Hume's view we have no other notion of cause and effect but that of certain objects, which have been always conjoined. Bertrand Russell adds that we cannot penetrate into the reason of the conjunction (*Russell on Hume*).

Although practical biomedical sciences tend generally not to be concerned during day-to-day work about causation with such philosophical rigour, the principles of causation must be taken into consideration in all work that examines interactions between events with regard to a possible cause and effect relationship. Francis Goulton, the inventor of finger-printing and cousin of Charles Darwin, measured the length of a person's arm and the size of that person's head and asked to what degree can one of these qualities predict the other. Goulton's experimental work, while technically simple, documented for the first time in the history of science that the correlation between two biological variables can be connected on the possible basis of measurement rather than human judgements, i.e. that the attribute to the correlation to variation of the two organs is partly due to common causes. This led Pearson

to state, 30 years later, that his interpretation of Goulton's work was that there was a category broader than causation, namely correlation, of which causation was only the limit. This new concept of correlation brought psychology, anthropology, medicine and sociology in large parts into the field of mathematical treatment. (Pearl, 2001)

In general, the two fundamental questions about causality are:

1. What experimental evidence is required for legitimate inference of a cause and effect relationship?
2. Given that we are willing to accept causal information about the phenomena, what inferences can we draw from such information and how?

Pharmacovigilance and pharmacoepidemiology are new scientific disciplines. In common with other biomedical sciences, causation is much harder to ascertain than correlation in these disciplines. There are many examples in biomedical sciences where correlations are generally accepted without full ascertainment of causality. For example, whilst science continues to work to identify the precise pathway of causation, to a patient and to society it is clear that the relationship between smoking cigarettes and lung cancer is accepted. Similarly, the relationship between the use of aspirin and the development of Reye's syndrome in children is generally accepted without full understanding of the pathophysiological mechanisms involved.

The notions of necessary and sufficient causes

The notions of necessary cause and sufficient cause are of interest in a number of applications, both in science and law. For example, oxygen is necessary for fire to set in a room; however, alone, oxygen is not sufficient for starting a fire and another action, such as striking a match, is needed to start a fire. This is very important in pharmacovigilance. For example, a person with partially patent coronary arteries may have no symptoms at rest but will get angina when exercising, with the development of severe anaemia or when receiving a medication that constricts the coronary arteries. Apart from very extreme levels, these three factors will not cause angina in a patient with fully patent coronary arteries. Therefore, the partial patency of the coronary arteries is a necessary but insufficient condition for the angina to occur. Since medicines are not always taken by people who are otherwise healthy, some people by the nature of their genetic constitution (which may affect the pharmacokinetic or pharmacodynamic responses to drugs) and patients with a previous medical history or other predisposing factors are more likely to experience some adverse drug reactions (ADRs) than people without such conditions. Awareness of the impact of such factors on the likelihood of developing specific ADRs is an important consideration in causality assessment.

Factors to be considered in causality assessment

In day-to-day medical practice, the boundaries between the concepts of a cause and a correlation are blurred; the word 'cause' is usually used somewhat loosely. This may have been driven by practical needs, since one of the reasons why patients consult doctors is to

know what has caused their condition. Many patients are disappointed if the doctor does not provide them with a possible cause. For example, a patient with a rash may be understandably disappointed if his doctor tells him that he is unsure whether the cause of the rash is a viral illness or an allergy. For practical clinical purposes, lack of precision in the application of the concept of 'cause' may lead to no harm as long as there is awareness that, in most cases, the interaction between the presumed cause and effect is a relationship, which in some cases may be a strong relationship, but not a definite cause. Although this notion is true to some extent when dealing with ADRs, the distinction between a cause and correlation needs to be more precise. For example, as nearly every patient will get bradycardia with high doses of beta-blockers and since bradycardia is a common adverse reaction with therapeutic doses of beta-blockers, it is not unreasonable to say that beta-blockers cause bradycardia in many patients. Conversely, while it is accepted that prolonged use of hormone replacement therapy (HRT) is associated with slight increase in the incidence of breast cancer (Dixon, 2001), and therefore a small number of long-term users will develop this condition, it is more appropriate to describe the interaction as an association. The reason for the difference between beta-blockers and HRT is the difference in the level of the certainty and strength of the association, both of which are much greater for the former than the latter. These two examples are relatively easy to handle, because of the abundance of good quality relevant data; but this is not the case for many other adverse events with medicines.

The clinical diagnosis of suspected ADRs is not different from the clinical diagnosis of conditions caused by other environmental factors, such as microbial or physical agents. The diagnostic process follows the usual steps of obtaining history, conducting an appropriate physical examination, ordering appropriate investigations and sometimes using time as a diagnostic tool. Some 30 years ago Irey (1972) described comprehensively the diagnostic problems in drug-induced disease and proposed strategies to aid the diagnoses of these conditions. Much of what was proposed by Irey remains applicable. For example, the emphasis on time to onset, the differential diagnoses (of causes other than the drug), the selection of the responsible drug on the basis of the pattern of the event or by exclusion, and the emphasis on rechallenge and dechallenge are all aspects that remain valid and applicable.

Performing causality assessment in pharmacovigilance may involve making a decision based on the information in a single adverse event or suspected ADR report (or a series of such reports) on the relationship between exposure to a drug and the reported adverse event or the suspected ADR. While this process is generally referred to as causality assessment, it is in fact an assessment of possible relatedness between exposure to a medicine and an adverse event. This evaluation which is useful in pharmacovigilance and essential in clinical medicine is different from the broader process of assessing causality of adverse reactions on the basis of data from multiple sources (see p. 338). There are many similarities between the clinical diagnosis of suspected ADRs in individual patients and causality assessment of case reports in pharmacovigilance. The points that need to be considered in both include:

1. The temporal relationship (time to onset)

The temporal relationship is perhaps the most important point to consider in assessing the relationship between exposure to a drug and a subsequent adverse event. Since most ADRs are 'Type A' (pharmacologically related) reactions, a plausible temporal relationship between exposure and the onset of the suspected ADR, taking into consideration the

pharmacological characteristics of the drug, is normally the first point to consider in the assessment of causality. A causal relationship is supported when the onset of an ADR coincides with the expected peak tissue concentration of the drug. Conversely, in most cases doubt should be raised when the onset of an ADR bears no relationship to the pharmacological characteristics of the drug.

With regard to measurement of the time to onset, it is important to consider the nature of the reaction. For short-term reactions, e.g. flushing with a calcium channel blocker, the relevant time to onset is the time between the last dose and the onset of the reaction. With long-term reactions, e.g. cataracts with systemic corticosteroids, the relevant period is the time between starting treatment and the development of the reaction.

2. The clinical and pathological characteristics of the events

Since ADRs are caused by definite pathophysiological mechanisms (although many are unclear at the time of assessment), conformity with the recognized clinical and pathophysiological patterns of the ADR is supportive of a causal relationship. For example, the clinical manifestations and laboratory abnormalities of patients with anaemia following the use of α -methyl dopa are expected to be consistent with the diagnosis of haemolytic anaemia (a recognized association) and doubt about relatedness should be raised when other types of anaemia are attributed to the use of this product.

3. Pharmacological plausibility

Pharmacological plausibility is based on previous knowledge of the drug, including its pharmacodynamic and pharmacokinetic characteristics. It is self-evident that causality is supported for events that can be explained on the basis of the pharmacological characteristics of the suspect drug, e.g. when the occurrence of the event coincides with the expected peak concentration in the affected organ.

4. Existing information

Whether the event has been previously reported as an adverse reaction (in clinical trials and post-marketing). Although the Summary of Product Characteristics (SmPC), or the Investigator Brochure for products in development, includes a list of adverse events or suspected ADRs that have been previously reported, the search for important events may need to include prescribing and pharmacological information in other sources.

5. Concomitant medication

The role of concomitant medications and medications taken prior to the event that could have caused the event. In patients who are taking more than one drug, it is frequently difficult to decide which one is the more likely cause of the suspected ADR. All the points that need to be considered for causality assessment for the suspect drug apply to concomitant medications or drugs to which the patient was exposed.

6. Underlying and concurrent illnesses

Some events which are attributed to drug exposure may be simply manifestations of pre-existing conditions. Furthermore, patients with some diseases, e.g. AIDS, respond differently to drugs, with qualitatively and quantitatively different ADR profiles. Patients with some conditions may experience altered responses to medications, e.g. patients with heart failure may experience augmented responses to cardiovascular medications.

7. Dechallenge or dose reduction

Recovery after stopping the drug or dose reduction is important, particularly when the timing of the full recovery or improvement is consistent with the pharmacological characteristics of the drug. However, it is noteworthy that some events, e.g. deafness with aminoglycosides, are irreversible.

8. Rechallenge or dose increase

The occurrence of ADRs after dose increases or rechallenge is a strong indicator of causality, particularly if the onset of the suspected ADR is pharmacologically plausible. A note of caution is to resist temptation for using rechallenge as a diagnostic tool unless clinically warranted and informed consent has been obtained. There must be a valid reason for the rechallenge and the practitioner and the institute must be trained and prepared for all eventualities, e.g. severe anaphylaxis and cardiac arrest.

9. Patients' characteristics and previous medical history

Previous medical history including previous history of drug allergies and the presence of renal or hepatic impairment can be very relevant in the diagnosis of suspected ADRs. For example, patients with atopy or a history of allergy to drugs are more likely to experience allergic ADRs. Female sex, low body weight, renal impairment and hepatic impairment are well known risk factors for ADRs (see Chapter 2).

10. Drug interactions

A plausible temporal relationship with the introduction or cessation of a concomitant drug with potential for interaction is an important consideration in causality assessment.

11. The quality of information in the report

Whereas minimal information is required for valid regulatory reports, causality assessment requires all the necessary data elements. In many cases in pharmacovigilance practice this can be obtained only after seeking detailed follow-up information from the reporter.

The clinical diagnosis of ADRs and causality assessment in pharmacovigilance require expertise that includes sound biomedical education and the ability to apply relevant clinical and pharmacological knowledge. Specifically, they require:

- Awareness that the event could be an ADR for one of the drugs that the patient is receiving. Therefore, practitioners and assessors should be aware of the common and serious ADRs for the medicines used, including their clinical patterns and other biological aspects.
- Access and ability to use information sources and they need to establish links with medical information departments or pharmaceutical companies for information on rare events or drugs with which they are unfamiliar.
- Basic knowledge of the pharmacokinetic (absorption, distribution, metabolism and excretion) and pharmacodynamic characteristics of the drugs used, as well as potential for drug interactions.

- Familiarity with special investigations that aid the diagnosis or monitoring of suspected ADRs, including their indications and limitations. Such investigations include measurement of drug blood levels, tissue biopsy, etc.
- Clinicians who are responsible for the management of patients with suspected ADRs must be trained to handle unclear situations. For example, if even after careful consideration it remains unclear which of several drugs that the patient is receiving is the most likely cause of a suspected ADR, then a strategy for stopping the drugs one at a time and evaluating the response to dechallenge may be needed. This requires weighing the benefits gained from each drug with the detrimental effects of its withdrawal and the management of the consequences of these clinical decisions.

Methods for causality assessment

Many systems for causality assessment have been proposed. These have been mostly used during regulatory reporting by pharmaceutical companies, assessment by regulatory authorities or in academic research. According to Jones (2000), these systems can broadly be divided into three categories:

1. Unrestricted clinical evaluation/global introspection.
2. Algorithms, with or without scoring.
3. Bayesian probabilistic methods.

Each of these categories has advantages and disadvantages, as discussed below.

Unrestricted evaluation/global introspection

This is the most commonly used method by clinicians and by many pharmaceutical companies and regulatory authorities. Although the method is easy to apply and resembles the day-to-day medical practice with its 'common-sense' approach, it lacks transparency and is subject to the fallibility of human judgement. The method permits inconsistency, as well as missing or misinterpreting important points.

Structured algorithms with or without scoring

There are many structured or semi-structured algorithms for causality assessment. These range from simple flow charts (with or without scoring) to lengthy questionnaires with tens of questions that need to be analysed by powerful computers (Jones, 2000). The advantages of structured assessments include transparency and consistency. When a structured method for causality assessment is applied properly, no points are missed. A consistent method of assessment facilitates communication between healthcare professionals and aids documentation and audit. Disadvantages include reduced ability to apply clinical judgement and, with algorithms that include scoring, adherence to the scores in place of clinical judgement.

With regard to algorithms, one that has gained popularity in pharmacovigilance assessment is the system proposed by Naranjo *et al.* (1981), perhaps because of its simplicity;

see Table 7.1. At a clinical level, Sackett (1991) proposed a scoring system for deciding whether a treatment had caused an adverse effect in a particular patient that is more comprehensive than the system proposed by Naranjo *et al.* (1981).

Table 7.1 A scoring strategy for deciding whether this treatment caused that adverse effect in this particular patient (Naranjo *et al.*, 1981). Reproduced by permission of Elsevier Science (Mosby)

	Yes	No	Don't know
1. Are there previous <i>conclusive</i> reports on this reaction?	+1	0	0
2. Did the adverse event appear after the suspected drug was administered?	+2	-1	0
3. Did the adverse reaction improve when the drug was discontinued or a <i>specific</i> antagonist was administered?	+1	0	0
4. Did the adverse reaction reappear when the drug was readministered?	+2	-1	0
5. Are there alternative causes (other than the drug) that could on their own have caused the reaction?	-1	+2	0
6. Did the reaction reappear when a placebo was given?	-1	+1	0
7. Was the drug detected in the blood (or other fluids) in concentrations known to be toxic?	+1	0	0
8. Was the reaction more severe when the dose was increased, or less severe when the dose was decreased?	+1	0	0
9. Did the patient have a similar reaction to the same or similar drugs in <i>any</i> previous exposure?	+1	0	0
10. Was the adverse event confirmed by any objective evidence?	+1	0	0

Categorization of causality (relatedness) for individual suspected adverse drug reaction reports

There have been many classification systems proposed for causality (Stephens, 1988). The numbers of categories in these systems range from six (certain, probably, possible, unlikely, conditional/unclassified and unassessable/unclassifiable) in the World Health Organization (WHO) system (Edwards and Biriell, 1994) to three in the European A,B,O system (Meyboom, 1998). In general, the systems use modifications of the factors initially proposed by Irey (1972). The Drug Safety Research Unit (DSRU) in Southampton uses a modified version of the WHO system. The system used by the DSRU includes the five categories defined below, viz. probable, possible, unlikely, awaiting further information and unassessable. The only difference between the DSRU and the WHO systems is that the former does not include the category 'certain', because it is very unusual to be certain about causality from a single case report and the term probable is considered sufficient for events that may be categorized as certain in the WHO system.

1. *Probable*: a clinical event, including a laboratory test abnormality, that is well defined clinically and/or pathologically, occurring with a reasonable time sequence to administration of the drug, more likely to be attributed to the drug than to concurrent disease or other drugs and which follows a clinically reasonable response on withdrawal or dose reduction (dechallenge) or reintroduction of the drug (rechallenge) or dose increment.

Other supporting criteria for a probable case include laboratory investigations, such as high drug levels, or histological findings supportive of a drug-induced effect.

2. *Possible*: a clinical event, including a laboratory test abnormality, with a reasonable clinical and pathological definition (not a vague description of a clinical entity), occurring with a reasonable time sequence to administration of the drug but which could also be explained by concurrent disease or other drugs or chemicals. Information on other criteria, such as dechallenge, rechallenge and confirmatory investigations, may be either not fully available or inconclusive.

Medical judgement is necessary to classify reports for which it is possible that the event was caused by the drug but another cause, e.g. an underlying illness is far more likely. These should normally be classified as 'unlikely'.

3. *Unlikely*: a clinical event, including a laboratory test abnormality, with a temporal relationship to drug administration that makes a causal relationship improbable or in which other drugs or underlying or concurrent diseases provide far more plausible explanations.
4. *Awaiting further information*: a clinical event, including a laboratory test abnormality, about which more data are essential for proper assessment or the additional data are under examination.
5. *Unassessable*: A report for which it was not possible to obtain the additional information necessary for an appropriate evaluation.

An important feature of post-marketing safety evaluation is the large number of reports, which even after follow-up remain with incomplete (sometimes with very limited) information. However, even with such limitations it is possible to assess individual reports to the five categories outlined above.

Bayesian probabilistic methods

In simple terms, the Bayesian approach to statistics is based on assigning a prior probability to the event under investigation. In drug safety, the prior probability of an adverse reaction is based on information obtained from pre-marketing clinical trials and epidemiological studies for patients with the underlying illness. That probability is then modified in the light of the information obtained from the new information. The revised probability is called the posterior probability (Kirkwood, 1994).

There has been long-standing academic interest in the application of Bayesian approaches to causality assessment (Jones, 2000), but the progress has not been as fast as the proponents of the methods would have hoped, nor have the methods been generally embraced by regulatory authorities or pharmaceutical companies. There are two main reasons for this:

1. While the Bayesian approach to statistics resembles, to some extent, human intelligence, the apparent complexity of its mathematics deters those unfamiliar with statistics.

2. Even with increasing availability of epidemiological and pharmacoepidemiological databases, background information for calculating prior probabilities is often either sparse or unavailable. Furthermore, pre-marketing data from clinical trials are mostly confidential between the pharmaceutical companies and regulatory authorities and are not available for the scientific community to be used for estimating the prior probability. Even when they are available, they may not be in a format that could be used for Bayesian evaluation of the causality of ADRs.

It is important that the Bayesian methods for causality assessment in pharmacovigilance continue to develop, because they could play an important role in the field.

When to assess causality (relatedness)

1. By the prescribing doctor or investigator for clinical research subjects.
The clinical diagnosis of ADRs, which is usually performed in an unstructured way, is an exercise of causality assessment. With regard to clinical research, doctors participating in clinical trials are required, according to study protocols, to conduct causality assessment for all adverse events regardless of causation.
The advantage of clinicians conducting causality assessment is that the practising doctor or the investigator has detailed knowledge of the patient's current condition(s), previous history, concomitant medications, etc., as well access to clinical documents and reports of investigations. The disadvantage is that clinicians may have not been trained to assess causality for ADRs and may apply an incomplete, incorrect or inconsistent approach.
2. When the report is initially received by the regulatory authority or a pharmaceutical company.
Early assessment of causality aids classification of reports for processing, e.g. whether the report should be sent to regulatory authorities in an expedited manner. The disadvantage of such early assessment is that initial reports may include limited information and further follow-up information is required for proper evaluation. However, from time to time a need may arise to conduct causality assessment for medically important, serious and expected reports to assess their impact on the risk–benefit profile of the product. However, it is extremely rare that public health decisions are based on the information included in a single case report.
3. When follow-up information is received.
The advantage is that causality assessment benefits from the maximum information that can be realistically obtained for an individual report. In nearly all circumstances the enhancement of assessment by the additional information compensates for the delay required to wait until all the necessary information for the evaluation is available.
4. At the time of signal generation.
5. At the time of investigating a safety issue or writing a periodic safety update report (PSUR).

Assessing causality from multiple information sources: the application of the Bradford-Hill criteria in pharmacovigilance and pharmacoepidemiology

Although the assessment of individual case reports or clusters of case reports is an important part of pharmacovigilance, drug safety signals and hypotheses are generated from several sources. In addition to spontaneous reports of suspected adverse reactions and published case reports, information sources for pharmacovigilance include animal studies, clinical pharmacology studies, clinical trials and pharmacoepidemiological studies. In an ideal world all generated signals would be evaluated and investigated further if necessary. Currently, this happens infrequently and usually in a limited way. Regulatory and clinical decisions are made on the data available at the time of evaluation. In 1998, the CIOMS IV Working Group raised concerns regarding the limitations and inconsistencies of safety evaluations by regulatory authorities of important drug safety issues such as product withdrawals (CIOMS, 1998). The CIOMS IV report provided pragmatic, yet comprehensive, and useful guidelines for safety evaluation. Together with earlier initiatives, this provides a good foundation to build on to improve safety evaluation, but the field requires further methodology, policy research and audit.

Although the safety of medicines must consider some specific issues, such as variation in compliance and drug interactions, the general principles that are used to study environmental hazards are applicable with some modifications. One of the most important papers published in the 20th century with thoughts on the epidemiological basis of disease causation was a summary of a lecture given by Sir Austin Bradford-Hill (who was then Emeritus Professor of Statistics at the University of London) entitled 'The environment and disease: association or causation' in 1965 (Bradford-Hill, 1965). In the introduction, he asked two simple questions: 'How in the first place do we detect the relationship between sickness, injury and conditions of work? How do we determine what is a physical, chemical and psychological hazard of occupation, and in particular those that are rare and not easily recognized?' Bradford-Hill described 'aspects' of an association that need to be considered before deciding on the most likely interpretation of its causation.

These aspects, listed in Table 7.2, are commonly referred to as the Bradford-Hill criteria for a causal association, and they have been used by epidemiologists and others when addressing causation of disease in a broad range of situations (Rothman and Greenland, 1998). In his lecture, Bradford-Hill described strengths and weaknesses of each of these aspects with regard to their contribution to an inference of causation. These criteria have

Table 7.2 The Bradford-Hill criteria

1.	Strength
2.	Consistency
3.	Specificity
4.	Temporality
5.	Biological gradient
6.	Plausibility
7.	Coherence
8.	Experimental evidence
9.	Analogy

been used to interpret evidence from pharmacoepidemiological studies. In the broader discipline of pharmacovigilance, consideration of these criteria, with modifications dictated by the nature of the data, can be very helpful in the interpretation of evidence from various sources at different levels. Shakir and Layton (2002) outlined thoughts on the application of Bradford-Hill's criteria to the evaluation of pharmacovigilance data with some examples.

Strength

Bradford-Hill stated that strong associations are more likely to be causal than weak associations. Weak associations are more likely to be explained by unrelated biases. For example, the association between smoking and lung cancer is so strong (studies show relative risks ranging between 10 and 30) that, even if some biases were operating, a shift of the association to non-causal is unlikely (Strom, 2000). In general epidemiology, a relative risk of <2 is considered to indicate a weak association (Strom, 2000). This is one of the major problems of pharmacoepidemiology, it is uncommon to find high relative risks (>2) for ADRs, particularly for serious ADRs, with marketed medicines, compared with placebo or with other products. In general, this is because medicines associated with a high incidence of serious ADRs would have been considered too toxic for marketing.

With regard to signal generation in pharmacovigilance, to complement qualitative methods, new approaches have been proposed such as proportional reporting rates (PRRs) (Evans *et al.*, 2001) and Bayesian confidence propagation neural network (BCPNN) (Bate *et al.*, 1998). In essence, these methods compare the proportion of a particular event reported for a drug with the whole database (which includes all or some of the other products monitored by the organization). A high PRR or information component (IC) in BCPNN suggests that an event has been reported more frequently than expected with the product and may indicate a safety signal. Although higher PRRs and ICs do not necessarily indicate a greater likelihood (Evans *et al.*, 2001) of causality, they do suggest stronger signals. It is possible to apply the Bradford-Hill strength criterion in signal generation in pharmacovigilance; for example, in a recent study using the WHO database of spontaneous reports of ADRs, BCPNN was used to examine the strength of the signals relating to heart muscle disorders with antipsychotic drugs (Coulter, 2001). Clozapine, an antipsychotic that had previously been reported to be associated with heart muscle disorders, had a much higher IC than lithium, a drug not known to be associated with such disorders. This example demonstrates the feasibility of applying the strength criterion in assessing the results of the quantitative methods for signal generation in pharmacovigilance.

Consistency

Bradford-Hill stated that repeated observations of an association in different populations under different circumstances provide additional support for a causal association. However, he cautioned that lack of consistency does not rule out a causal association, because some effects may be produced by the causes only in certain circumstances.

Because of the low relative risks of ADRs generally detected in pharmacovigilance and pharmacoepidemiological studies, the consistency of findings in different populations is highly important. For example, in studies conducted to examine the association between the use of the commonly called 'third-generation' oral contraceptives (OCs; gestodene and desogestrel) and deep-vein thrombosis, although the strength of the association in some of

the studies may have been weak by conventional epidemiological standards (a relative risk between 1.5 and 2) (Hannaford, 2000), the consistency of the finding of a higher risk among users of the third-generation OCs in different populations utilizing different methods supports an inference of causation.

With regard to pharmacovigilance studies, reporting of a particular event in different populations is supportive of a true association. An example is the association between the antiepileptic, lamotrigine, and serious skin reactions (Stevens–Johnson syndrome and toxic epidermal necrolysis); this was strengthened by the fact that spontaneous reports were sent both from hospitals and the communities in several countries, and that cases were reported in clinical trials (Guberman *et al.*, 1999) and in a prescription-event monitoring (PEM) study (Mackay *et al.*, 1997).

The utilization of the Bradford-Hill consistency criterion (by conducting several studies with different methodologies in different healthcare systems to study the safety of a particular drug) is perhaps the most important approach to address any uncertainties resulting from the low relative risks that are commonly detected in pharmacoepidemiological studies.

Specificity

Bradford-Hill stated that a cause leads to a single effect, not multiple effects; but he cautioned that, although the concept of specificity is sometimes useful, it could be misleading.

In pharmacovigilance, specificity is important because drugs cause ADRs by specific mechanisms, which may or may not be known at the time of the enquiry. Associations between the use of some drugs and an increase in the incidence of cancer have been reported. For example, prolonged use of HRT is associated with a slight increase in the incidence breast cancer (Dixon, 2001). True associations such as this one are specific. In most circumstances it is not plausible that the use of a drug is associated with an increase in the incidence of multiple cancers. In the 1990s, a controversy was raised by a suggestion of an association between the use of intramuscular vitamin K in neonates and childhood cancer (Golding *et al.*, 1992). It is difficult to think of a mechanism by which a single injection of vitamin K can lead to increasing the likelihood of development of several cancers in childhood. Subsequent studies were more focused in studying the association between injectable vitamin K in neonates and specific cancers, e.g. acute lymphoblastic leukaemia (Ansell *et al.*, 1996). This demonstrates the need for specificity when considering a causal association in pharmacoepidemiology and pharmacovigilance.

Because of the incomplete information in many spontaneous reports, attention must be paid, by trying to obtain follow-up information, to ensure that the reports include sufficient information to enable a specific diagnosis. For example, if a drug is suspected to be associated with anaemia, then confirmation needs to be sought regarding the type of anaemia and whether there is consistency between the reported cases clinically and haematologically (with regard to the type of anaemia).

Temporality

Bradford-Hill stated that there is a necessity that cause precedes effect in time. This is very important in pharmacovigilance. Many ADRs occur in patterns based on relation to exposure, the pharmacological characteristics of the drug and host responses (Type A

reactions). A consistent pattern with regard to temporal relationship is, therefore, very important in assessing the causal relationship between drug exposure and an adverse event or a cluster of events. The need for a plausible and generally consistent temporal relationship also applies to Type B (unpredictable) ADRs. Conversely, although an inconsistent pattern does not exclude a causal relationship, one should be doubtful of a possible causal relationship.

Studying temporal relationships in pharmacovigilance should take into account the effects of drug interactions (e.g. when the pharmacological plausibility that an ADR is related to a drug is supported by the timing of the introduction or stopping a concomitant medication) or the occurrence of illnesses (e.g. renal impairment) or a physiological state (e.g. dehydration).

Biological gradient

Bradford-Hill stated that a biological gradient as demonstrated by a dose–response curve is well known in epidemiology. It has been demonstrated in numerous studies that the number of cigarettes smoked and the number of years of smoking are directly related to the development of major smoking-related hazards (cancer of the lung and cardiovascular diseases) (Strom, 2000). However, care should be taken, because biological gradient can sometimes be complex. For example, although it is well known that excessive drinking of alcohol is associated with detrimental dose-related effects, drinking small amounts of alcohol can be protective (Burns *et al.*, 2001).

In pharmacovigilance, a causal association is supported when an ADR occurs in a dose-dependent manner, or from cumulative exposure over a prolonged period of time. There are many examples when causal association was supported by plausible dose relationships, such as the oestrogen content of the combined OC pills and deep-vein thrombosis (Inman *et al.*, 1970) and the systemic effects of nasal and inhaled corticosteroids (CSM/MCA, 1998).

Plausibility

Bradford-Hill stated that biological plausibility of a hypothesis is another aspect to be considered for causal inference. He added that plausibility is an important concern that may be difficult to judge.

In pharmacovigilance, plausibility is easy when the mechanism is known, as in the case of non-steroidal anti-inflammatory drugs (NSAIDs) and gastrointestinal bleeding. However, it is difficult when the mechanism is unknown, as for many Type B ADRs. In such situations, an ADR may not be detected readily, e.g. the delay in the recognition of the association between the use of practolol and the oculomucocutaneous syndrome (Wright, 1975).

Research is needed to study whether the strength of plausibility can influence the levels of relative risk needed to accept an inference of causation, i.e. whether it possible with strong pharmacological plausibility to accept a lower level of relative risk as an indicator of causal association than when the plausibility is weak.

Coherence

Bradford-Hill defined coherence as the cause-and-effect interpretation whose data should not seriously conflict with generally known facts of the natural history and biology of a disease. The principle of coherence is useful in pharmacovigilance and pharmacoepidemiol-

ogy. For example, the proposed possible association between a single intramuscular injection of vitamin K, the neonatal period and childhood cancers lacked coherence with the understanding at the time of the pathogenesis of cancer (Golding *et al.*, 1992). However, one has to keep an open mind regarding associations that are not coherent with contemporary knowledge, because understanding of the biological basis of ADRs often comes at a later stage.

Experimental evidence

Experimental evidence as a supporter of causal inference is self-evident. Studies in biological models, as well as in animal and human experiments, may all lend support to signals raised by pharmacovigilance. It is commonly necessary to conduct experimental studies to better understand the signals generated in pharmacovigilance and pharmacoepidemiology, e.g. when a drug interaction that was not previously evaluated is suspected.

Analogy

Bradford-Hill said that inventive scientists could find analogies everywhere. An analogy finds a source of more elaborate hypothesis about an association under study. As elsewhere in biomedical sciences, analogies can guide or mislead. For example, while cough has been reported in 5–20 per cent of patients who take angiotensin converting enzyme (ACE) inhibitors, the association between cough and angiotensin II (A_{II}) receptor antagonists, e.g. losartan and irbesartan, has no pharmacological plausibility (Hardman and Limbird, 2001). Therefore, in evaluating reports of cough with A_{II} antagonists, an analogy with ACE inhibitors is inappropriate.

Conversely, the analogy between COX-2 inhibitors and conventional NSAIDs with regard to the gastrointestinal safety profile continues to be appropriate until we fully understand the interaction between the pharmacological selectivity of the COX-2 inhibitors and their clinical effects.

Conclusion

The application of Austin Bradford-Hill's criteria for evaluating causal associations in pharmacovigilance and pharmacoepidemiology is very useful. This requires understanding of the general characteristics of pharmacovigilance data, e.g. under-reporting, misclassification and poor quality of information from third parties. Further work is required to propose ways to handle such limitations.

References

- Ansell P, Bull D and Roman E (1996). Childhood leukaemia and intramuscular vitamin K: findings from a case-control study. *Br Med J* **313**: 204–205.
- Bate A, Lindquist M, Edwards IR, Olsson S *et al.* (1998). A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* **54**(4): 315–321.
- Bradford-Hill A (1965). The environment and disease: association or causation. *Proc R Soc Med* **58**: 295–330.

- Burns J, Crozier A and Lean ME (2001). Alcohol consumption and mortality: is wine different from other alcoholic beverages? *Nutr Metab Cardiovasc Dis* **11**(4): 249–258.
- CIOMS (1998). *Benefit risk balance for monitored drugs: evaluating safety signals*. Report of CIOMS Working Group IV, Geneva.
- Coulter DM, Bak A, Meyboom RHB, Lindquist M, Edwards R (2001). Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study. *BMJ* **322**: 1207–1209.
- CSM/MCA (1998). The safety of inhaled and nasal corticosteroids. *Curr Probl Pharmacovig* **24**: 6.
- Dixon JM (2001). (Editorial) Hormone replacement therapy and the breast. We should worry about the increase in the risk of breast cancer. *Br Med J* **323**: 1381–1382.
- Edwards IR and Biriell C (1994). Harmonisation in pharmacovigilance. *Drug Saf* **10**: 93–102.
- Evans SJW, Waller PC and Davis S (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* **10**: 483–486.
- Golding J, Greenwood R, Birmingham K and Mott M (1992). Childhood cancer, intramuscular vitamin K, and pethidine given during labour. *Br Med J* **305**: 341–346.
- Guberman AH, Besag FM, Brodie MJ, Dooley JM *et al.* (1999). Lamotrigine-associated rash: risk/benefit considerations in adults and children. *Epilepsia* **40**: 985–991.
- Hannafor P (2000). Cardiovascular events associated with different combined oral contraceptives: a review of current data. *Drug Saf* **22**: 361–371.
- Hardman JG and Limbird LE (eds). (2001). *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 10th edition. McGraw-Hill: 694–696, 828, 833.
- Inman WH, Vessey MP, Westerholm B and Engelund A (1970). Thromboembolic disease and the steroidal content of oral contraceptives. A report to the Committee on Safety of Drugs. *Br Med J* **ii**: 203–209.
- Irey NS (1972). Diagnostic problems in drug-induced diseases. In *Drug-induced Diseases*, vol. 4, Meyler L and Peck HM (eds). Excerpta Medica: Amsterdam; 1–24.
- Jones JK (2000). Determining causation from case reports. In *Pharmacoepidemiology*, Strom BL (ed.). Wiley: Chichester; 525–538.
- Kirkwood B (1994). *Essentials of Medical Statistics*. Blackwell Sciences: 73.
- Mackay FJ, Wilton LV, Pearce G, Freemantle SN *et al.* (1997). Safety of long term lamotrigine in epilepsy. *Epilepsia* **38**: 881–886.
- Meyboom RHB (1998). *Detecting Adverse Drugs Reactions*. The Netherlands Pharmacovigilance Foundation, LAREB, 's Hertogenbosch; 87–88.
- Naranjo CA, Busto U, Sellers EM, Sandor P *et al.* (1981). A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* **30**(2): 239–245.
- Pearl, J. (2001) *Causality reasoning and inference*, Cambridge University Press pp. xiii–xvi.
- Rothman KJ and Greenland S (eds) (1998). *Modern Epidemiology*, 2nd edition. Lippincott Williams & Wilkins: 24–28.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991). *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd Edition. Lippincott-Raven: Philadelphia: 285.
- Shakir SAW and Layton D (2002). Causal association in pharmacovigilance and pharmacoepidemiology. *Drug Saf* **25**(6): 467–471.
- Stephens MDB (1988). The diagnosis of adverse medical events associated with drug treatment. In *The Detection of New Adverse Drug Reactions*, 2nd edition. Macmillan: 73–114.
- Strom BL (2000). *Pharmacoepidemiology*, 3rd edition. Wiley: Chichester.
- Wright P (1975). Untoward effects associated with practolol administration: oculomucocutaneous syndrome. *Br Med J* **i**: 595–598.