REVIEW ARTICLE

Drug Safety 2008; 31 (1): 21-37
0114-5916/08/0001-0021/$48.00/0

Methods for Causality Assessment of Adverse Drug Reactions

A Systematic Review

Taofikat B. Agbabiaka,[1] *Jelena Savović*[2] and *Edzard Ernst*[1]

1 Complementary Medicine, Peninsula Medical School, Universities of Exeter and Plymouth, Exeter, UK
2 Department of Social Medicine, University of Bristol, Bristol, UK

Contents

Abstract

Numerous methods for causality assessment of adverse drug reactions (ADRs) have been published. The aim of this review is to provide an overview of these methods and discuss their strengths and weaknesses. We conducted electronic searches in MEDLINE (via PubMed), EMBASE and the Cochrane databases to find all assessment methods. Thirty-four different methods were found, falling into three broad categories: expert judgement/global introspection, algorithms and probabilistic methods (Bayesian approaches). Expert judgements are individual assessments based on previous knowledge and experience in the field using no standardized tool to arrive at conclusions regarding causality. Algorithms are sets of specific questions with associated scores for calculating the likelihood of a cause-effect relationship. Bayesian approaches use specific findings in a case to transform the prior estimate of probability into a posterior estimate of probability of drug causation. The prior probability is calculated from epidemiological information and the posterior probability combines this background information with the evidence in the individual case to come up with an estimate of causation. As a result of problems of reproducibility and validity, no single method is universally accepted. Different causality categories are adopted in each method, and the categories are assessed using different criteria. Because assessment methods are also not entirely devoid of individual judgements, inter-rater reliability can be low. In conclusion, there is still no method universally accepted for causality assessment of ADRs.

Adverse drug reactions (ADRs) are frequent major causes of morbidity, hospital admissions and even death.[1-4] Causality assessment is the evaluation of the likelihood that a particular treatment is the cause of an observed adverse event.[5] It assesses the relationship between a drug treatment and the

occurrence of an adverse event. It is an important component of pharmacovigilance, contributing to better evaluation of the risk-benefit profiles of medicines[6] and is an essential part of evaluating ADR reports in early warning systems and for regulatory purposes.[7] The idea of creating a standardized assessment for the relationship-likelihood of case reports of suspected ADRs was conceived in the hope that this would, in a structured way, lead to a reliable and reproducible measurement of causality.

In an early seminal paper[8] written from the perspective of occupational exposure, the author defined nine aspects of association between environmental factors and disease that should be considered in order to arrive at a decision regarding causation. These aspects are the *strength* of the association, *consistency* of the observed association, *specificity*, *temporality* of the association, the *biological gradient* (dose-response curve), *plausibility*, *coherence*, *experiment* and *analogy*. Although the paper does not specifically discuss adverse drug reactions it is nevertheless relevant to the process of assigning causality.

Establishing causal association between suspected drugs and subsequent clinical events is obviously important. Causality assessment methods differ in many respects but share certain common features. To arrive at a conclusion on the 'suspect' drug, questions are used to capture details of the ADR. Different procedures are thereafter adopted to convert answers from these questions to estimates of probability.[9] Causality assessment methods vary as to the criteria and categories used to assign causality (tables I and II).

The spontaneous reporting system in pharmacovigilance is a process of collecting, assessing, presenting and interpreting suspected ADRs. Case reports acquired from spontaneous reports are assessed by first evaluating cases individually and secondly interpreting the aggregated data.[43,44] This evaluation is a broader way of evaluating data from multiple sources, although similar in some respects to clinical diagnosis in individual patients. Causality assessment in pharmacovigilance may involve making a decision based on the information on the relationship between a drug exposure and suspected ADR from a single adverse event or suspected ADR report (or a series of reports). Standardized causality assessment is now a routine procedure at pharmacovigilance centres around the world; it is aimed at decreasing ambiguity of the data and also plays a key role in data exchange and limits the drawing of erroneous conclusions.[43]

Various methods have been published for assessing causality in ADRs, but because there are no defined diagnostic criteria or categories, inter-rater and intra-rater variability can be large.[17,45,46] Currently, there is no universally accepted method for assessing causality of ADRs. No up-to-date review of the existing causality assessment methods is available. This review aims to fill this gap.

To identify published methods, scales and instruments for causality assessment of ADRs, electronic literature searches were conducted in the following electronic databases: MEDLINE (via PubMed), EMBASE and the Cochrane Library, all from inception to April 2007. The search terms used were 'adverse event OR side effect OR adverse drug reaction OR adverse drug event' and 'causality OR causal OR algorithm'. The MeSH® (Medical Subject Heading) descriptors 'adverse drug reaction' OR 'adverse event' and 'causality' were also searched on MEDLINE. No language restriction was applied to the search or inclusion of articles.

Further relevant publications were located by manually searching reference lists of retrieved articles.

Only publications relating to instruments, methods or scales for causality assessment of individual ADR reports in humans are included in this review. The review does not include methods used in causality assessment of aggregated ADR reports. We also excluded studies on adverse reactions from agents other than drugs, such as therapeutic failures and intentional and accidental poisoning.[5] For the purpose of this review, drugs included synthetic herbal medicines, vitamins and other supplements.

All articles deemed eligible for inclusion based on titles and abstracts were retrieved and read in full. Data relating to causality assessment method, ADR or adverse drug event definitions, criteria and categories for assigning causality, causality categories adopted, and results from testing the merits and limitations of the methods or assessment scales were recorded.

Table I. Published methods for causality assessment of adverse drug reactions (ADRs): criteria for assigning causality

Author	TTO/ temp seq	Prev exp/ drug info	Alt aetiol cands	Drug level / evidence of OD	Chall-enge	De-chall	Re-chall	Response pattern to drug	Confirmed by lab evidence	Con-comitant drugs	Background epidemiol /clin info	ADR char/ mech	Other
Expert Judgement or Global Introspection													
Wilholm[10]	✓	✓	✓	✓	×	×	✓	✓	x	✓	x	×	×
WHO & UMC[5]	✓	×	×	✓	×	×	✓	×	✓	×	×	×	✓
Miremont et al.[11]	✓	×	×	×	×	×	✓	✓	✓	✓	×	✓	×
Arimone et al.[7]	✓	×	×	✓	×	×	✓	✓	✓	×	×	×	✓
Algorithms													
Irey[12]	✓	×	×	✓	x	✓	✓	✓	✓	×	×	×	×
Karch and Lasagna[13]	✓	✓	✓	×	×	✓	✓	×	×	×	×	×	×
Dangoumau et al.[14]	×	×	✓	✓	×	✓	✓	✓	×	×	×	×	×
Begaud et al.[15]	×	×	✓	✓	×	✓	✓	✓	×	×	×	×	×
Kramer et al.[16]	✓	×	✓	✓	✓	×	✓	✓	×	×	×	×	×
Blanc et al.[17]	✓	×	×	×	✓	×	×	×	✓	×	×	×	×
Emanueli and Sacchetti[18]	✓	✓	✓	×	✓	×	✓	✓	×	×	×	×	×
Naranjo et al.[19]	×	✓	✓	×	✓	×	✓	✓	✓	×	×	×	×
Jones[20]	×	✓	×	✓	×	×	✓	✓	×	×	×	×	×
Evreux et al.[21]	✓	✓	✓	×	×	✓	×	×	×	×	×	×	×
Kitaguchi et al.[22]	✓	×	✓	×	×	✓	×	✓	×	×	×	×	×
Lagier et al.[23]	×	×	×	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
Cornelli[24]	✓	×	×	✓	×	×	✓	✓	×	✓	×	×	×
Stephens[25]	×	×	✓	✓	×	×	✓	✓	✓	✓	×	×	×
Castle[26]	✓	×	✓	✓	×	×	×	✓	×	✓	✓	×	×
Turner[27]	✓	×	×	×	×	✓	✓	×	×	×	✓	×	×
Venulet et al.[28]	×	✓	×	✓	×	✓	✓	×	×	✓	×	×	✓
Loupi et al.[29]	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
Stricker et al.[30]	×	✓	×	✓	×	×	×	×	✓	×	×	×	×
Benichou and Danan[31]	✓	×	×	×	×	×	×	✓	✓	×	✓	×	×
Hoskins and Mannino[32]	✓	×	×	×	×	✓	✓	×	×	×	✓	✓	×

Continued next page

Table I. Contd

Author	TTO/ temp seq	Prev exp/ drug info	Alt aetiol cands	Drug level / evidence of OD	Chall-enge	De-chall	Re-chall	Response pattern to drug	Confirmed by lab evidence	Con-comitant drugs	Background epidemiol /clin info	ADR char/ mech	Other
Danan and Benichou[33]	✓	×	✓	×	×	×	×	✓	✓	×	✓	×	×
Hsu and Stoll[34]	×	✓	×	✓	×	×	✓	✓	×	×	×	×	×
Maria and Victorino[35]	✓	✓	✓	✓	×	×	×	✓	✓	×	×	✓	×
Koh and Shu[36]	✓	×	✓	✓	✓	×	✓	×	×	×	×	×	×
Horn et al.[37]	✓	✓	✓	×	×	✓	✓	✓	✓	×	✓	✓	✓
Probabilistic/Bayesian Approaches													
Marshford[38]	✓	×	×	×	×	×	×	×	✓	×	×	×	✓
Lane et al.[39]	✓	×	×	×	×	×	✓	✓	✓	×	×	✓	×
Hutchinson et al.[40,41]	✓	×	×	×	×	×	✓	✓	✓	✓	×	×	×
Lanctot et al.[42]	✓	✓	×	×	×	✓	✓	×	×	×	✓	✓	×

ADR char/mech = characteristics/mechanisms of ADR; **Alt aetiol cands** = alternative aetiological candidates; **Background epidemiol/clin info** = background epidemiological or clinical information; **Dechall** = dechallenge; **OD** = overdose; **Other** = other factors; **Prev exp/drug info** = previous experience or information on drug; **Rechall** = rechallenge; **Temp seq** = temporal sequence; **TTO** = time to onset.

Our search yielded a total of 2695 references, 95 of which were considered potentially relevant after reading the abstract (figure 1). A total of 34 papers met our inclusion criteria: 25 from the electronic searches and 9 from reference lists. The methods, scales and instruments fell into three main categories: (i) opinion of experts, clinical judgement or global introspection (n = 4); (ii) algorithm or standardized assessment method (n = 26); and (iii) probabilistic or Bayesian approaches (n = 4).

1. Expert Judgement or Global Introspection

Identifying ADRs often depends on the clinical assessments of physicians, usually the clinician treating the patient or, at other times, a clinical pharmacologist.[32] It is a process whereby an expert expresses judgement about possible drug causation by considering all available data relevant to a suspected ADR,[47] estimates their relative importance and assigns weights to deduce the probability of the role of the drug in the untoward event.[32] Many reports of ADRs are based on the judgement of a single evaluator; others carry out assessments using a group of experts, or experts and non-experts, who then compare their evaluations to arrive at a consensus-based conclusion. Such judgements are based on knowledge and experience, but even among experts, frequent disagreements occur.[48] We identified four published methods based on expert opinion or global introspection (tables I and II).

The assessment method used by the Swedish regulatory agency is based on expert judgements.[10] The clinician evaluates the causal relationship by considering seven different factors: (i) the temporal sequence, (ii) previous information on the drug, (iii) dose relationship, (iv) response pattern to drug, (v) rechallenge, (vi) alternative aetiological candidates and (vii) concomitant drugs. Events are classified as 'probable' or 'possible' and 'non-assessable' or 'unlikely'. A limitation of this method is the small number of categories into which causality can be placed, as there may be an overlap and ADRs could be wrongly evaluated.[10]

A review of the advances and limitations of causality assessment by the WHO and Uppsala Monitoring Centre (UMC), in consultation with the national centres who participated in the WHO Pro-

Table II. Published methods for causality assessment of adverse drug reactions: causality categories

Author	Prob/ likely	Causative	Def	Poss	Coincidental	Exclude	Unclassified/ conditional	Doubtful	Remote/ unlikely	Unassessable/ unclassifiable	Certain	Unrelated	Neg
Expert Judgement or Global Introspection													
Wilholm[10]a	✓	×	×	✓	×	×	×	×	✓	✓	×	×	×
WHO and Uppsala Monitoring Centre[5]	✓	×	×	✓	×	×	✓	×	✓	✓	✓	×	×
Miremont et al.[11]	✓	×	✓	×	×	×	×	✓	✓	×	×	×	×
Arimone et al.[7]	✓	×	×	✓	×	✓	×	✓	✓	✓	✓	×	×
Algorithms													
Irey[12]	✓	✓	×	✓	✓	×	×	×	×	×	×	×	✓
Karch and Lasagna[13]	✓	×	✓	✓	×	×	✓	×	×	×	×	✓	×
Dangoumau et al.[14]	✓	×	×	✓	×	×	×	✓	×	×	×	×	×
Begaud et al.[15]	✓	×	×	✓	×	×	×	✓	×	×	×	×	×
Kramer et al.[16]	✓	×	✓	✓	×	×	×	×	✓	×	×	✓	×
Blanc et al.[17]	✓	×	×	✓	✓	×	×	✓	×	×	✓	×	×
Emanueli and Sacchetti[18]	✓	×	✓	✓	×	✓	×	✓	×	×	×	×	×
Naranjo et al.[19]	✓	×	✓	✓	×	×	×	✓	×	×	×	×	×
Jones[20]	✓	×	×	✓	×	×	×	×	✓	×	×	×	×
Evreux et al.[21]	×	×	×	×	×	×	×	✓	×	×	✓	×	✓
Kitaguchi et al.[22]	✓	×	✓	✓	×	×	×	×	✓	×	×	×	×
Lagier et al.[23] b													
Cornelli[24]	✓	×	✓	✓	×	×	×	✓	×	✓	×	×	×
Stephens[25]	✓	×	×	✓	×	×	×	×	✓	×	✓	×	×
Castle[26] b													
Turner[27]	✓	×	×	✓	×	×	×	×	✓	×	×	✓	×
Venulet et al.[28]	✓	×	✓	✓	×	×	×	×	×	✓	×	✓	×
Loupi et al.[29]	×	×	×	×	×	×	×	✓	×	×	✓	×	×
Stricker et al.[30]	✓	×	✓	✓	×	✓	×	×	✓	×	×	✓	×
Benichou and Danan[31]	✓	×	×	✓	×	×	✓	×	×	×	×	✓	×
Hoskins and Mannino[32] b													

Continued next page

Table II. Contd

Author	Prob/ likely	Causative	Def	Poss	Coincidental	Exclude	Unclassified/ conditional	Doubtful	Remote/ unlikely	Unassessable/ unclassifiable	Certain	Unrelated	Neg
Danan and Benichou[33]	✓	×	×	✓	×	✓	×	×	✓	×	×	×	×
Hsu and Stoll[34]	✓	×	✓	✓	×	×	×	×	×	×	×	×	×
Maria and Victorino[35]	✓	×	✓	✓	×	✓	×	×	✓	×	×	×	×
Koh and Shu[36]	✓	×	✓	✓	×	×	×	×	✓	×	×	×	×
Horn et al.[37]	✓	×	×	✓	×	×	×	✓	×	×	×	×	×
Probabilistic/Bayesian Approaches													
Marshford[38]	✓	×	×	✓	×	✓	×	×	✓	×	✓	×	×
Lane et al.[39]	✓	×	×	×	×	✓	×	×	×	×	✓	×	×
Hutchinson et al.[40,41] [b]													
Lanctot et al.[42] [b]													

a Events were classified into one of two categories: 'probable/likely or possible' and 'remote, unlikely or unaccessable/unclassifiable'.

b Causality categories not indicated.

Def = definite; **Neg** = negative; **Poss** = possible; **Prob** = probable.

gramme for International Drug Monitoring, led to the development of a practical tool for case report assessment.[5] The method takes into consideration the clinical-pharmacological aspects of the case history and the quality of the observation. Causality is grouped into one of six categories based on a number of assessment criteria. The method gives guidelines on selection of different categories of ADR. Defined criteria are also used to exclude events. Previous knowledge and statistical chance are not considered in this method.

Miremont et al.[11] compared physicians' opinions on drug-event relationship with the scores obtained by the French causality assessment method (discussed in further detail in section 2).[49] Physicians assigned very high scores (i.e. 'very likely' or 'likely') to 60% of the events assessed and low scores (i.e. 'unlikely', 'dubious/possible') to 32% of the cases. The French causality assessment method, on the other hand, scored 89% of the cases low (i.e. as 'dubious/possible') and only 11% as 'likely'. Both methods only agreed in 6% of the cases assessed. A conclusion from this study was that physicians tend to assign very high scores to suspected ADRs.

Another method[7] had five experts independently assess cause-effect relationship on a random set of ADRs. Their judgements were expressed on a 100mm visual analogue scale (VAS), ranging from 0 to 1 probability of causation. These probabilities were further split into seven levels of causality (see tables I and II) for easier analysis. Results from the 30 ADR cases evaluated showed that probabilities given on the VAS differed amongst the five experts; they only agreed on the same level in 17.3% of cases. Level of agreement varied according to the level of causality: moderate (kappa coefficient [κ] = 0.40) for 'exclude', low for 'likely' (κ = 0.32) and 'certain' (κ = 0.30). Overall inter-rater agreement was poor (κ = 0.20). Even after causes of discrepancies had been discussed and a consensus reached, a high degree of disagreement was still observed.

2. Algorithms

An algorithm is a problem-specific flow chart with step-by-step instruction on how to arrive at an answer.[50] It is a clinical instrument in the form of a questionnaire that gives detailed operational criteria

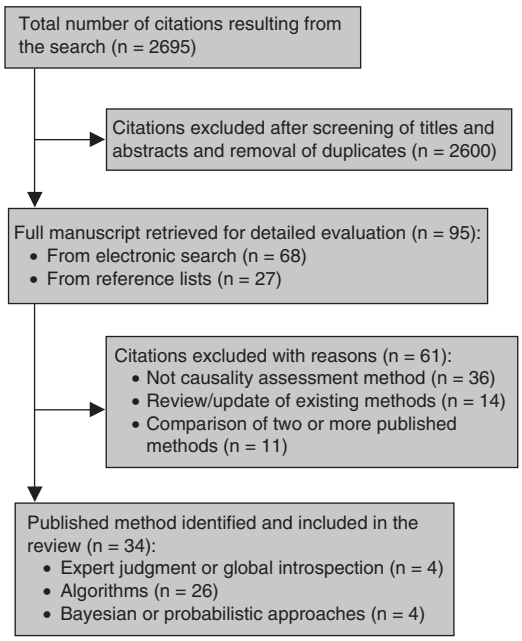

Fig. 1. Flow chart of identified, excluded and included studies.

for ranking the probability of causation when an ADR is suspected.[51] Algorithms give structured and standardized methods of assessment in a systematic approach to identifying ADRs based on parameters such as time to onset of the ADR or temporal sequence, previous drug/adverse reaction history and dechallenge and rechallenge. Individual cases are approached systematically, resulting in a high degree of consistency and reproducibility.[52] Clinical judgement is, however, required at various stages to arrive at a conclusion.[53,54] The majority of algorithms share basic common features in the form of a set of questions enabling the user to gather adequate information about the case in order to arrive at an objective conclusion.[55] Various algorithms have been developed and published to resolve issues of bias, reproducibility and validity in causality assessment.[6,45] However, no single algorithm is accepted as the 'gold standard', because of the shortcomings and disagreements that exist between them, as highlighted in other publications.[26,45]

The first algorithm to be developed considered six criteria that may link a drug with an event and classified five degrees of causality.[12] A drug is considered eligible for causation if fully identified and administered within a reasonable period before an event occurred. The weakness of this well described method lies in its emphasis on pathological data at the expense of clinical data; therefore, assessing causality may be difficult or almost impossible in the absence of pathological information.

Karch and Lasagna[13] developed a decision-table approach to identify potential ADRs. The three tables respectively identify potential drug reactions, assess the certainty of the link between the drug and the event and evaluate the underlying causes of the identified untoward events. Diagnosis of ADRs using this method requires judgements about alternative aetiological candidates, timing, previous experience with the drug in question, and information on dechallenge and rechallenge. The advantage of this method lies in the fact that the decision tables are usually more compact than other algorithms and easier to use manually, and it can equally be adapted to computer language using a decision logic translator. However, this algorithm still requires the operator to make certain important judgements or refer to medical texts in evaluating diagnostic evidence to arrive at conclusions on causality. Thus, it is not strictly an operational procedure; therefore, results may not always be reproducible. This algorithm is also unable to identify new ADRs or first cases of ADRs, as it requires previous bibliographical description of the adverse event. Also, individual cases may not be adequately evaluated since a case classified as 'probable' may be categorized as 'definite' by another evaluator.

The method now regarded as the 'French method' has been used by the French regulatory agency since 1977. It is based on a three-stage process: assessment of chronological criteria, clinical and biological findings and assessment of symptoms for the detection of ADRs.[14] The method separates an intrinsic imputability (possible cause between drug and clinical event) from an extrinsic imputability (bibliographical data) using seven criteria (three chronological and four semiological) in two different tables. The chronological criteria are (i) drug challenge, (ii) dechallenge and (iii) rechallenge, with an overall score of four possible categories. The semiological criteria are (i) semiology (clinical signs) *per se* (suggestive or other), (ii)

favouring factor, (iii) alternative non-drug-related explanation (none or possible) and (iv) specific laboratory test with three possible outcomes (positive, negative or no test for the event-drug pair). Total score from these four semiological scores are grouped as semiological imputability with three possibilities ('likely', 'possible' and 'dubious'). When combined, scores from these two decision tables (chronological and semiological) give an intrinsic causality of five possible categories.[49] The advantage of this method is that it allows certain drugs taken at the same time with the 'suspect' drug to be excluded, because each drug is imputed separately. However, this method requires more time than most other algorithms. Begaud et al.[15] published an actualization of the French method[14] with the same principles of separating intrinsic imputability from extrinsic imputability.

Kramer et al.[16] expanded on a previous algorithm[13] to develop a new set of criteria with specific rules for ADR assessment. This algorithm applies to a single clinical manifestation occurring after administration of a single suspect drug. In cases where multiple drugs are involved, each is assessed separately. The algorithm is made up of six decision tables with a scoring system incorporated into each axis. A score of +1, 0 or −1 is assigned for the weight of evidence on each axis. Individual scores from each of the axes are added to obtain a total score, which corresponds to overall probability of an ADR. Cases are assigned to probability categories based on a total score that ranges from −7 to +7. Individual judgement is required at each stage in order to arrive at a decision on the clinical evidence. Therefore, clinical manifestations, markers or indicators following the administration of a drug need to be carefully characterized by the observer. The algorithm is useful in cases where more than one drug is suspected. With a little modification, it can be used to assess drug interactions or cases where withdrawal rather than administration is likely to be responsible for observed clinical manifestations. This algorithm was tested among four practicing clinicians and four interns. It had an impressive effect on interobserver agreements between the senior clinicians (33% without algorithm to 77% with algorithm, weighted κ = 0.27–0.67) but did not significantly affect agreement levels between interns.[51] The authors suggested that the need for a clinical judgement in order to arrive at final conclusions may be the reason for the non-significant agreement observed when the algorithm was used by the less experienced interns. One of the advantages of this algorithm is its transparency. The reasons for disagreement between observers can be identified and decisional steps where mistakes were made can be re-examined and corrected. However, certain levels of expertise, experience and time are required to use this method effectively.

Another decision table was designed by Blanc et al.[17] to assess the nature of the relationship between the drug and the event. It considers three factors: time sequence, response pattern and role of underlying disease(s). The time sequence has four categories (C_1, C_1a, C_2 and C_3). The response pattern has four categories (S_1, S_1a, S_2 and S_3) and the role of underlying disease has three categories (M_1, M_2 and M_3). Combining these three decisional axes gave five causality levels ranging from 'doubtful' to 'certain' (see table II). Low concordance was observed using this algorithm; only 25% of ADRs evaluated were attributed to the same drug by all of the three observers.[17]

Emanueli and Sacchetti[18] developed an algorithm in the form of a decision table for the analysis of ADR reports in large clinical trials, based on the concepts of an earlier algorithm.[13] Events are classified on a 5-point scale according to their relationship with the suspect drug. The algorithm has eight questions that are answered either 'yes' or 'no'. This algorithm is an improvement on earlier methods as it replaces the 'drug related or not' concept with a 5-point scale probability level of linkage between the drug and the observed reaction. The method is equally adequate and of practical use in large scale trials. However, it is difficult to rate events higher than 'possible' using this method if other causes such as clinical state or other therapies cannot be ruled out.

Another ADR probability scale consisting of ten questions that are answered as 'yes', 'no', 'unknown' or 'inapplicable' was developed by Naranjo et al.[19] to assess causality in a variety of clinical situations using the conventional categories and definitions of 'definite', 'probable', 'possible' and 'doubtful'. Scores assigned to each question ranged

from −1 to +2. The event is assigned to a probability category based on the total score. A total score of ≥9 is 'definite', 'probable' is 5–8, 'possible' 1–4 and 'doubtful' ≤0. This scale is intended to assess the likelihood of an ADR associated with only one drug, not for adverse drug events resulting from interactions between two drugs.[56] The Naranjo scale does not address the main points that are necessary in causality evaluation of potential drug interactions. Nevertheless, the adverse reaction scores obtained by using the Naranjo scale correlate well with those derived with Kramer's algorithm, which does address those points.[16]

A review of existing algorithms by the US FDA in 1982 led to the development of a relatively simple algorithm for their own assessment of ADRs.[20] Based on the general concepts of previous methods,[12,13] cases were analyzed using four characteristics leading to three causality categories. Although very quick and simple to use, this method does not consider previous evidence related to the drug. The four criteria on which causality is judged may also not clearly capture causality, as there may be an overlap of one category with another. The categories to which causality are assigned are also not discriminatory. Therefore, the method may only be useful for specific ADR assessments, since it may be difficult to make conclusions on clear-cut degrees of causality.

Evreux et al.[21] proposed a method for the assessment of drug-induced adverse reaction reports by imputability according to the report's importance in support of aetiology. Criteria for assigning causality include drug-related factors such as previous analogous events, published information, age (availability) of the drug on the market. Associated (non-drug dependent) factors such as environment and traumatic events are also taken into consideration. Events are categorized as 'certain', 'doubtful/questionable' or 'absent'. Assessing ADEs using this method may be more difficult since there are numerous factors to be considered.

A post-marketing system to assess the relationship between a suspected drug and an ADR, proposed by Kitaguchi et al.,[22] is based on a scrupulous analysis of some data on individual patients who experienced the ADR. A series of analogous cases are assessed in terms of the time relationship, reproducibility and specificity of the association between drug and event. Six questions, answered as 'yes', 'unknown' or 'no', are used to capture information for cause-effect assessment of the suspected ADR.

The method known as the 'balanced assessment method'[23] evaluates case reports on a series of VASs, according to the likelihood that each criterion is fulfilled. The nine causality criteria (see table I) considered in this method were grouped into three categories. Each factor is judged to be either caused by the drug or some other causes and rated 0 to 1 on the analogue scale (0 representing an alternative cause and 1 representing a drug as the cause). One of the merits of this method is that it considers the possibility of an alternative to causation for each of the factors and not just as a separate factor. Although each case is assessed by two independent assessors, the evaluation still depends, to a large extent, on the level of assessor's knowledge. An evaluator needs to be an expert in the particular area to make reliable evaluations. Computation of relative risk in this method is similar to the prior odds ratio in the Bayes' theorem discussed in section 3 and consideration of alternative causes for each factor also resembles estimating likelihood ratios.

Another algorithm considered five criteria to assign causality into six different categories based on the overall score. Each of the five criteria is scored between 1 and 4; hence the maximum score is 20. An overall score of <11 is categorized as 'unrelated', while a score of ≥16 is considered 'definite'. This method is fast and easy to use, but previous knowledge of the drug event or dechallenge is not considered in the assessment.[24]

An algorithm to assess ADR causality in both pre- and post-marketing surveillance and to distinguish between several drugs using a wide range of scores was published in 1984.[25] The method is based on scoring six criteria, which correspond to the criteria previously used in other algorithms.[16,19] Scores range between 1 (denoting that a causal relationship is unlikely) to 5 or 6 (causality link is neutral) and 10 (indicating a highly likely link). The method is useful in separating concurrent drugs and disease from the event in assigning causality.

A 'summary time plot' rather than a set of questions was proposed for the industrial setting to iden-

tify patterns of ADRs.[26] The plot summarizes the time relationship between treatment and a possible adverse reaction. With sufficient information from the causality criteria, the duration of treatment and possible adverse reaction are plotted with time on the x-axis and severity of the possible adverse reaction on the y-axis. This method does not lead to a conclusion on causality, because it only summarizes the time factor alongside other factors that are relevant to the drug-event relationship. The method is, however, quick to use, saves the use of ambiguous terminology and is applicable even with minimal information.

A modified version of the Jones algorithm[20] for ADR cases resulting in death was published by the FDA's division of Drug Experience.[27] Causality assessment using this algorithm is based on four basic principles: temporal eligibility, dechallenge, rechallenge and confounding factors. Although this algorithm allows for the amount of information in any case report to be determined or classified, only the basic information needed for making causal relationship assessments is provided. A lack of reference to previous reports about the drug or event and relevant literature in the ADR assessment is a major limitation of this method.

The 'Ciba-Geigy method' resulted from a number of expert consensus meetings. Experts used their clinical judgement to assess events and assign causality on a VAS.[28] This method was updated[57] and replaced with a checklist of 23 questions, split into three sections: (i) history of present adverse reaction, (ii) patient's past adverse-reaction history and (iii) monitoring-physician's experience. Each question was assigned numerical values in multiples of five. A degree of causality was assigned according to the total score.[57] This updated method was found to have a high degree of agreement (62%) when compared with evaluators' assessments. Although the level of reliability does not assure validity, this method reflects the knowledge and experience of the evaluator, and the type of ADR that is being evaluated.

A simple algorithm made up of two sections and three main questions was developed to assess teratogenicity, as no previous method existed for this condition.[29] The first sections of the algorithm (chrono-semiological axis) allow for the drug to be excluded if not implicated in the origin of the abnormality. The second section (bibliographical axis) weights the bibliographical data. The three questions consider alternative aetiological candidates other than the drug, chronology of the suspect drug and other bibliographical data, to arrive at a conclusion on causality.

A diagnostic decision tree with a complementing algorithm developed by Stricker et al.[30] is a step-by-step guide to a detailed assessment of suspected liver injury events. The evaluation is completed by using the algorithm to assess the cause-effect relationship. Three factors – temporal relationship between drug and hepatic event, the course of the event and other laboratory evidence – are considered while other possible causes are excluded in order to make a conclusion on causality.

A decision support algorithm aided by integrated informatics has been developed to structure and organize decision processes in ADR assessment.[32] This algorithm is based on an analytical hierarchical decision-making process in the software package CRITERIUM© (Sygenex, Redmond, WA, USA). Causality criteria are rated relative to each other on a consistency scale and combined with individual judgements to obtain an overall rating that may be either a probability or a likelihood scale. The merit in this method lies in the fact that multiple causes or multiple drugs are considered in the final analysis with no special knowledge of statistics required for its operation. The package is interactive, giving feedback at every stage of the analysis. However, decision making using this instrument may be jeopardized by unorganized and low-quality data. The best available data and good conceptual understanding of the decision-making process and the medical condition assessed are required to arrive at a reliable conclusion.

Following the outcome of an International Consensus Meeting[58] of experts on definitions and causality criteria of ADRs, a new scale and weighted scores, the Roussel Uclaf Causality Assessment Method (RUCAM), was proposed for predetermined disease states such as liver and dermatological injuries.[33] Causality criteria were assigned weighted values with the total sum giving a numerical score. Each criterion had a scale from −3 to +3 corresponding to the role of the drug evaluated. This

method was expected to produce better reproducibility in evaluations than previous methods. This is because disagreement between observers had been explained and the scores for each criteria agreed by the panel of experts. A retrospective assessment[33] of the reproducibility of this method among four experts showed a 37–99% agreement rate. Although this method seems quite easy to use, it is organ-specific. Therefore, the criteria need to be defined by a consensus of experts for each medical field and validated before it can be of any meaningful use in ADRs other than hepatic or dermatological injuries.

Another algorithm to guide routine drug causality assessment in clinical trials was published in 1993.[34] Known associations between drug and event were used to modify causality scores based on results of dechallenge and rechallenge. This algorithm is similar to that developed by the FDA described earlier in this section.[20] The method prompts the evaluator to identify all possible aetiologies in the event rather than just searching for concordance with the drug-causation hypothesis. This algorithm is limited by the fact that, although dechallenge may give an indication of causation when followed to its resolution, deliberate rechallenge is unethical in most situations.

Maria and Victorino[35] developed a scale known as the M&V scale for diagnosing drug-induced liver injury (DILI) based on seven causality criteria, as opposed to only three considered in an earlier proposed decision tree.[30] Probability was expressed as a score between −6 and 20, divided into five causality degrees. Diagnosis of DILI is complex and requires experienced clinicians in order to be accurate. The M&V clinical scale may improve assessments of the probability that a particular drug is involved in DILI. However, in cases where more than one drug is suspected, the scale needs to be computed for individual drugs. In such cases, points such as I-A, I-B and I-C of the scale distinguish two or more drugs. It therefore becomes difficult to discriminate between drugs with the same levels of probability. Some questions on the M&V scale apply only to immunoallergic hepatitis, making it difficult for scores to be generated for other hepatic injuries.

In 2005, another algorithm[36] was developed by matching, modifying and eliminating questions that cannot be answered from the information available on most ADR reporting forms. The 56 questions from a previous algorithm[16] were replaced with only eight questions and a scoring table in the new version.[36] Although this new algorithm is quick and easy to use since it does not require extra data other than those routinely collected on most ADR reporting forms, it is likely that other relevant information that could help evaluate causality is omitted. The lower threshold adopted in assigning cases as 'definite' may also result in more false-positive 'definite' cases, but in the long run, these could serve as early warning signals for use in drug monitoring.

Recently, Horn et al.[37] proposed the Drug Interaction Probability Scale (DIPS) to evaluate drug interaction cases. The DIPS uses ten questions that are answered 'yes' or 'no' to yield a score estimating the likelihood of drug interaction. Points are added for each affirmative response and subtracted for a negative response. The questions concern the pharmacological properties of the drug, the possible role of other drugs and specific patient information. The method was developed to assist users in the assessment of drug-interaction-induced adverse outcomes and also to serve as a guide for the further study of potential drug interactions. Although based on the Naranjo scale,[19] the DIPS is modified to reflect the differences between a single-drug event and one due to a drug-drug interaction. However, adequate knowledge of either the drugs involved and/or the basic mechanisms of interaction are necessary for accurate assessments. In addition, limited information on potential drug-interaction cases may result in low causation scores, leading to a false-negative evaluation of causation. Currently, the DIPS is not widely used, and therefore very little user feedback is available for its modification.

Tables I and II are summaries of the different causality criteria and categories adopted in published algorithms.

3. Probabilistic or Bayesian Approaches

Bayesian methods for causality assessment make use of specific findings in a case to transform a prior into a posterior probability of drug causation.[41] A series of likelihood ratios is also calculated for every relevant element of the case. A likelihood ratio (i.e. case-specific information, such as timing or rechal-

lenge that helps to distinguish between causes) is further broken down into components. Each component applies to a specific category of case information, and the final result is obtained by multiplying out the various terms to obtain a posterior probability of drug causation.[31] This method allows the simultaneous assessment of multiple causes. It is open-ended with no limit to the amount of case details that can be assessed.[40]

The Australian method for causality assessment was one of the first ADR assessment methods to be based on a probabilistic approach.[38] Conclusions are drawn from internal evidence, such as timing, and laboratory information from case reports. Previous knowledge on the suspect-drug profile is deliberately excluded in the assessment.[38] Probability decisions are made only on the likelihood of a causal relationship. This method differs from the Swedish method,[10] which relies to a large extent on previous knowledge about the drug to arrive at conclusions on causality.

A number of programmes for diagnosis of various ADRs using Bayes' theorem have been developed.[59,60] The Bayesian Adverse Reactions Diagnostic Instrument (BARDI) was developed to overcome the numerous limitations associated with expert judgements and algorithms.[60] It is a decisional analysis tool whereby probability of causation is calculated from epidemiological and clinical data (prior odds) and case analysis (likelihood ratio). The output is expressed as a posterior probability ranging from 0% (not drug induced) to 100% (definitely drug induced). The method is reproducible, and information on how posterior probability was obtained in various reactions can be obtained. It is explicit in the information used and how each piece of information is weighted.[38,39] The complex calculations involved in this method are one of its major limitations. It also requires more time and more expertise than most other assessment methods.

The Bayesian approach can be implemented as a spreadsheet programme on either paper or computer. It calculates and provides instant numerical and graphical feedback as soon as new pieces of evidence of the suspected ADR are evaluated.[40] Case reports are read and descriptions that fit reports from the literature are listed to help assess the prior probability. Elements to distinguish potential causes are also considered and noted. Scoring is done by assigning a score of 1 to the element in the list that is least likely to cause the event. The next lowest likely cause is assigned a score of between 1 and ∞. All potential causes listed in each category are compared with the potential cause that scored 1 in that category; scores under each category are then multiplied to obtain a final score. The software consists of worksheets to impute case parameters, one for case findings and another for scoring. Detailed step-by-step instructions on how to carry out the analysis are also provided. Although this method requires some expertise to operate, it can evaluate more than two possible causes at the same time. Individual assessments can also be incorporated into the probabilistic system to estimate ADR incidence based on data from post-marketing surveillance. The spreadsheet allows rapid calculations and interaction during the process.[31] It is not an easier method for causality assessment; it is, rather, a framework for a valid and internally consistent assessment.[61]

Computerization of the BARDI to reduce the number of errors, improve consistency and enhance the user's access to the system led to the development of the MacBARDI-Q&A.[50] MacBARDI is a prototype Bayesian-based computer programme developed as a clinical aid for the diagnosis of hypersensitivity rashes. It is a model of the original BARDI spreadsheet enhanced by a macro automating command so that less time is required to obtain information for input since the program has direct access to specific databases.[62] The MacBARDI-Q&A is user-friendly with improved interaction in a question-and-answer format when compared with the MacBARDI spreadsheet used alone. Other prototypes of BARDI developed for diagnosing ADRs include a model for the prediction of risk of pseudo-allergic reaction and histamine release in patients undergoing surgery[42] and a diagnostic aid for pseudomembranous colitis.[31] Some of the advantages of the BARDI include reliability (the same input information generates the same output), explicitness and transparency (final results show clearly what information is considered and its contribution in the assessment) and aetiological balancing (all drug and non-drug possible causes are considered in the assessment). The significant amount of time, resources and complex calculations

Table III. Comparisons of published methods for adverse drug reaction (ADR) causality assessment

Method	Comparison with other method/scale
WHO and Uppsala Monitoring Centre[5]	45.1% congruency[66] when compared with method by Kramer et al.[16]
Miremont et al.[11]	6% agreement compared with VAS[51]
Irey[12]	55% agreement; $\kappa = 0.33$ compared with GI[62]
Karch and Lasagna[13]	71% agreement compared with expert opinion[67] 41% agreement; $\kappa = 0.23$ compared with GI[62] More likely to assign a risk of 'unlikely' ADE ($p = 0.0001$)[53,54] than method by Kramer et al.[16]
Kramer et al.[16]	67% agreement among non-experts using method[16] 69% complete agreement among raters ($\kappa = 0.618$, $p < 0.001$) More likely to assign a possible risk of ADE ($p = 0.0001$)[53,54] than method by Karch and Lasagna[13]
Emanueli and Sacchetti[18]	59% agreement with Begaud;[49] $\kappa = 0.19$[62]
Naranjo et al.[19]	High correlation of ADR scores with Kramer et al.;[16] $r = 0.82$, $p < 0.001$[67] 65–67% agreement with Kramer et al.;[16] $\kappa_w = 0.43$–0.51[62] 38% agreement with Miremont et al.;[11] $\kappa = 0.34$[62]
Jones[20]	67% agreement with Kramer et al.;[16] $\kappa_w = 0.48$[66] 64% agreement with Naranjo et al.;[19] $\kappa_w = 0.28$[66] 40.2% congruency[66] with Kramer et al.[16] 47% agreement with GI; $\kappa = 0.24$[62]
Kitaguchi et al.[22]	56% agreement with GI; $\kappa = 0.36$[62]
Begaud[49]	59% agreement in computerized comparison with Jones;[20] $\kappa_w = 0.41$[62] 50% agreement with Kramer et al.;[16] $\kappa = 0.40$[62]
Stricker et al.[30]	Classified 66% of cases as 'probable' while Hsu and Stoll[34] distributed cases evenly among all categories
Danan and Benichou[33]	Good correlation with Hsu and Stoll[34] for drug-induced liver injury[66]
Hsu and Stoll[34]	49% agreement compared with GI; $\kappa = 0.24$[62]
Maria and Victorino[35]	Low absolute agreement of 18% with Kramer et al.;[16] $\kappa_w = 0.28$[66] High agreement of 84% compared with expert opinion; $\kappa = 0.93$[66]

ADE = adverse drug event; **GI** = global introspection; **VAS** = visual analogue scale; κ = kappa coefficicient; κ_w = weighted κ; **r** = correlation coefficient.

involved are obvious limitations of this approach.[62] The Bayesian approach is regarded as the most logical method for causality assessment and is recommended as a gold standard against which other methods should be tested.[63] Tables I and II are summaries of published probabilistic approaches for ADR causality assessment.

4. Perspective

Although expert judgement seems to be the most widely used causality assessment method, it fails to achieve the level of reproducibility and validity required of an assessment method. A number of publications evaluated the validity and reproducibility of this method, either by testing retrospectively on case reports, [7,46,64] or by comparison with other assessment methods (table III).[11] Expert judgement is often used alongside other assessment methods in an attempt to produce evaluations that are more valid. A few algorithms and some decision models for ADR assessment have been developed by experts' consensus. The model RUCAM, for determining causality in acute liver injury, was developed based on the results of consensus meetings on ADRs.[65] Similar scales were also proposed for other medical conditions, such as neutropenia, haemolytic anaemia and photosensitivity reactions.

Several reports have shown expert judgement not to be a reliable method for causality assessments in ADRs.[7,46,65] Although agreement in 80–83% of cases was reported in one comparison of an algorithm with expert opinion,[54] the method leaves room for preconceived opinions and subjectivity and does not guarantee a consistent approach. Expert opinions are uncalibrated and decision making is not explicit, transparent or reproducible.[68] Dif-

ferent conclusions are reached even when experts are presented with the same information about an ADR, because each expert has different experiences, knowledge and skills, views and biases that may affect the final conclusions. The main argument against this method is that it has poor reproducibility, inter-rater and intra-rater disagreements and no standardized clinical evaluations. These limitations notwithstanding, expert judgement is still popular in ADR assessments.[13,19,51,54] This may be because it is quite similar to clinical diagnosis, straightforward and easier to use than the other methods. However, the search for a method to overcome these limitations has led to the development of numerous algorithms and probabilistic methods.

Standardized methods yield better inter-rater reliability than expert judgement, and sources of disagreement can be identified.[19] Higher inter-rater agreement is found in algorithms for questions of a factual nature than for those requiring clinical judgements.[69] Inter-rater agreement improved substantially when algorithms were used rather than individual judgement.[51] However, intra-rater reproducibility is less frequently reported in the literature. This may be because it is considered unnecessary since inter-rater reliability demonstrates all sources of error contributing to intra-rater reproducibility as well as any other differences between observers. Yet, algorithms lack flexibility, and there is usually no room for the user to include additional information in the assessment. The method is based on finite information; therefore, data available from the case report, but not considered by the methodology, would have no influence on results.[52]

The probabilistic or Bayesian methods are considered more valid for causality assessment than expert opinion or algorithm.[70,71] The prototypes MacBARDI and MacBARDI-Q&A, developed for the computer-aided assessment of adverse drug events, are rapid, processing medical and pharmacological data without relying on clinical judgement alone. The complex and extensive calculations involved in this method, however, make it unpopular among clinicians. Computations of prior odds or likelihood ratios are other obvious drawbacks of this approach since the incidence and epidemiological data required for these calculations are often unavailable.[32] However, unlike algorithms, which are limited by finite information that can be evaluated, Bayesian approaches have no limits to the number of causal factors that can be evaluated, so multiple causes can be assessed simultaneously.[40]

Alternative criteria to assess reproducibility and validity in ADR assessment methods have been proposed.[9] These are repeatability, explicitness, transparency, completeness, aetiological balancing and the absence of *a priori* constraints on the effects of factors. The authors concluded that reproducibility and validity are inappropriate criteria for evaluating causality assessment methods, since reproducibility suppresses rather than resolves agreement among observers and validity relies on expert opinion, which itself is insufficiently reliable to be considered a gold standard. An evaluation of ten standardized causality assessment methods[13,17-20,23,25,28,49,54] showed that most of the methods failed in most of the criteria proposed by these authors.[9]

Different assessment methods adopt different criteria (table I) and categories (table II) for assigning causality. The definitions and causality categories currently endorsed by the WHO are not consistently used in majority of the ADR-assessment literature. These criteria have been criticized as being subjective and imprecise since they are based mainly on expert judgements.[17,46] Adopting equivocal and operational definitions in ADR studies is important. Failure to do so accounts for much of the inconsistency in causality evaluations. Some of the ADR scales (e.g. the WHO-UMC scale) do not have the 'unrelated' category that some investigators, particularly in clinical trials, find very important. The French causality assessment method[14] omitted extreme degrees such as 'definite' and 'non-drug related' because they consider the reliability of any method too poor to pronounce such categorical judgements. Some algorithms have three to four categories of causality,[16,19,20] while others have five or more.[17,31,72] It is therefore difficult to compare results from other algorithms with the WHO-UMC method, which has six categories.

The implicit use of definitions, inconsistent use of terminology and incomplete information from case reports coupled with vast differences in the application of clinical judgements are the major problems facing ADR evaluation.[72,73] The fact that

clinical judgements need to be applied at different steps in an algorithm may result in the inconsistent application of the algorithm to the same data by different users. Incomplete reporting equally makes it difficult to determine the role of concomitant drugs and other associated diseases in ADRs. Considering the fact that case reports are relied on for clinical decisions, drug policies and regulation, it is imperative that ADR reports meet some basic criteria. Adequate information on time to onset, rechallenge, confounding factors, dechallenge, degree and duration of exposure, background epidemiological and clinical information must be included in case reports.[3,74] These important data are often missing, but are, however, crucial for proper causality assessment.

In addition, confounding variables are recognized as a problem in establishing causality, especially with decision algorithms. The ability of algorithms to establish causality in ADRs is undermined by confounding variables decreasing the concordance of the result. When results from an algorithm were compared with those obtained from an expert panel, agreement was only on average 47%.[6] Agreement between global introspection and decisional algorithms varied according to levels of imputability. In the absence of confounders, agreement between the two methods was 41–69%, and 15–53% for reports with at least one confounding variable. At no level of causality assessment was full agreement found between decision algorithm and global introspection. Therefore, confounding variables compromised the sensitivity and specificity of algorithms.[45]

5. Conclusions

The numerous published methods for causality assessment in ADRs have various advantages and disadvantages. The idea of creating standardized causality assessment systems to provide reliable and reproducible measures of the relationship-likelihood in suspected cases of ADR seems unfeasible, since no single method has achieved this to date. The differences in ADR causality criteria and the unavoidable subjectivity of judgements may be responsible for the lack of reproducibility of most published methods. So far, no ADR causality assessment method has shown consistent and reproducible measurement of causality; therefore, no single method is universally accepted.

Acknowledgements

No sources of funding were used to assist in the preparation of this review. T.B. Agbabiaka and J. Savovic have both been supported by research fellowships sponsored by Dr Willmar Schwabe Pharmaceuticals, Germany. The authors have no conflicts of interest that are directly relevant to the contents of this review.

References

1. Schneeweiss S, Hasford J, Gottler M, et al. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. Eur J Clin Pharmacol 2002; 58: 285-91
2. Prince BS, Goetz CM, Rihn TL, et al. Drug-related emergency department visits and hospital admissions. Am J Hosp Pharm 1992; 49: 1696-700
3. Liu BA, Knowles SR, Mittmann N, et al. Reporting of fatal adverse drug reactions. Can J Clin Pharmacol 2001; 8: 84-8
4. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998; 279: 1200-5
5. World Health Organization (WHO), Uppsala Monitoring Centre. The use of the WHO-UMC system for standardized case causality assessment. WHO [online]. Available from URL: http://www.who-umc.org/graphics/4409.pdf [Accessed 2007 Oct 26]
6. Macedo AF, Marques FB, Ribeiro CF, et al. Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel. Pharmacoepidemiol Drug Saf 2005; 14: 885-90
7. Arimone Y, Begaud B, Miremont-Salame G, et al. Agreement of expert judgment in causality assessment of adverse drug reactions. Eur J Clin Pharmacol 2005; 61: 169-73
8. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965; 85: 295-300
9. Hutchinson TA, Lane DA. Assessing methods for causality assessment of suspected adverse drug reactions. J Clin Epidemiol 1989; 42: 5-16
10. Wiholm BE. The Swedish drug-event assessment methods. Special workshop – regulatory. Drug Inf J 1984; 18: 267-9
11. Miremont G, Haramburu F, Begaud B, et al. Adverse drug reactions: physicians' opinions versus a causality assessment method. Eur J Clin Pharmacol 1994; 46: 285-9
12. Irey NS. Tissue reactions to drugs. Am J Pathol 1976; 82: 613-48
13. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. Clin Pharmacol Ther 1977; 21: 247-54
14. Dangoumau J, Evreux JC, Jouglard J. Method for determination of undesirable effects of drugs [in French]. Therapie 1978; 33: 373-81
15. Begaud B, Evreux JC, Jouglard J, et al. Imputation of the unexpected or toxic effects of drugs: actualization of the method used in France. Therapie 1985; 40: 111-8
16. Kramer MS, Leventhal JM, Hutchinson TA, et al. An algorithm for the operational assessment of adverse drug reactions: I. Background, description, and instructions for use. JAMA 1979; 242: 623-32

17. Blanc S, Leuenberger P, Berger JP, et al. Judgments of trained observers on adverse drug reactions. Clin Pharmacol Ther 1979; 25: 493-8
18. Emanueli A, Sacchetti G. An algorithm for the classification of untoward events in large scale clinical trials. Agents Actions Suppl. 1980; 7: 318-22
19. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981; 30: 239-45
20. Jones JK. Adverse drug reactions in the community health setting: approaches to recognizing, counseling, and reporting. Fam Community Health 1982; 5: 58-67
21. Evreux JC, Loupi E, Descotes J, et al. Evaluation of the documentation on adverse effects of drugs: proposition 2. Imputability [in French]. Therapie 1982; 37: 657-70
22. Kitaguchi T, Nojiri T, Suzuki S, et al. Some assessment systems for industry post marketing adverse drug reaction (ADR) Information. Iyakuhin Kenkyu 1983; 14: 980-92
23. Lagier G, Vincens M, Castot A. Imputability in drug monitoring: principles of the balanced drug reaction assessment method and principal errors to avoid. Therapie 1983; 38: 303-18
24. Cornelli U. The phase IV monitoring studies: example of Naproxen Na multi-centre Italian trial. In: Crescioni C, James JM, editors. Postmarketing clinical trials – the phase V studies. Paris: Editions de Santé, 1984: 89-105
25. Stephens MD. Assessment of causality in industrial setting. Special workshop – industrial. Drug Inf J 1984; 18: 307-13
26. Castle WM. Assessment of causality in industrial settings. Drug Inf J 1984; 18: 297-302
27. Turner WM. The Food and Drug Administration algorithm. Special workshop – regulatory. Drug Inf J 1984; 18: 259-66
28. Venulet J, Ciucci A, Berneker GC. Standardised assessment of drug-adverse reaction associations: rationale and experience. Int J Clin Pharmacol Ther Toxicol 1980; 18: 381-8
29. Loupi E, Ponchon AC, Ventre JJ, et al. Imputability of a teratogenic effect [in French]. Therapie 1986; 41: 207-10
30. Stricker BHC. Diagnosis and causality assessment of drug-induced hepatic injury. In: Dukes MNG, ed. Drug-induced hepatic injury. Amsterdam: Elsevier, 1985: 1-13
31. Benichou C, Danan G. Causality assessment in the European pharmaceutical industry: presentation of the preliminary results of a new method. Drug Inf J 1992; 26: 589-92
32. Hoskins RE, Mannino S. Causality assessment of adverse drug reactions using decision support and informatics tools. Pharmacoepidemiol Drug Saf 1992; 1: 235-49
33. Danan G, Benichou C. Causality assessment of adverse reactions to drugs: I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. J Clin Epidemiol 1993; 46: 1323-30
34. Hsu PH, Stoll RW. Causality assessment of adverse events in clinical trials: II. An algorithm for drug causality assessment. Drug Inf J 1993; 27: 387-94
35. Maria VA, Victorino RM. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. Hepatology 1997; 26: 664-9
36. Koh Y, Shu CL. A new algorithm to identify the causality of adverse drug reactions. Drug Saf 2005; 28: 1159-61
37. Horn JR, Hansten PD, Chan LN. Proposal for a new tool to evaluate drug interaction cases. Ann Pharmacother 2007; 41: 674-80
38. Mashford ML. The Australian method of drug-event assessment. Special Workshop – regulatory. Drug Inf J 1984; 18: 271-3
39. Lane DA, Kramer MS, Hutchinson TA, et al. The causality assessment of adverse drug reactions using a Bayesian approach. Pharm Med 1987; 2: 265-83
40. Hutchinson TA, Dawid AP, Spiegelhalter DJ, et al. Computerized aids for probabilistic assessment of drug safety: I. A spreadsheet program. Drug Inf J 1991; 25: 29-39
41. Hutchinson TA. Computerized Bayesian ADE assessment. Drug Inf J 1991; 25: 235-41
42. Lanctot KL, Kwok MCO, Naranjo CA. Computerized Bayesian evaluation of adverse events. Drug Inf J 1995; 29: 319-25
43. Meyboom RHB, Hekster YA, Egberts ACG, et al. Causal or casual? The role of causality assessment in pharmacovigilance. Drug Saf 1997; 17: 374-89
44. Meyboom RHB. Causality assessment revisited. Pharmacoepidemiol Drug Saf 1998; 7 Suppl. 1: S63-5
45. Macedo AF, Marques FB, Ribeiro CF, et al. Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel, according to different levels of imputability. J Clin Pharm Ther 2003; 28: 137-43
46. Karch FE, Smith CL, Kerzner B, et al. Adverse drug reactions: a matter of opinion. Clin Pharmacol Ther 1976; 19: 489-92
47. Arimone Y, Begaud B, Miremont-Salame G, et al. A new method for assessing drug causation provided agreement with experts' judgment. J Clin Epidemiol 2006; 59: 308-14
48. Venulet J. Aspects of standardization as applied to the assessment of drug-event associations. Drug Inf J 1984; 18: 199-210
49. Begaud B. Standardized assessment of adverse drug reactions: the method used in France. Special workshop – clinical. Drug Inf J 1984; 18: 275-81
50. Lanctot KL, Naranjo CA. Computer-assisted evaluation of adverse events using a Bayesian approach. J Clin Pharmacol 1994; 34: 142-7
51. Leventhal JM, Hutchinson TA, Kramer MS, et al. An algorithm for the operational assessment of adverse drug reactions: III. Results of tests among clinicians. JAMA 1979; 242: 1991-4
52. Venulet J. Recognition and assessment of adverse drug reactions in children. Bratisl Lek Listy 1991; 92: 540-3
53. Frick PA, Cohen LG, Rovers JP. Algorithms used in adverse drug event reports: a comparative study. Ann of Pharmacother 1997; 31: 164-7
54. Hutchinson TA, Leventhal JM, Kramer MS, et al. An algorithm for the operational assessment of adverse drug reactions: II. Demonstration of reproducibility and validity. JAMA 1979; 242: 633-8
55. Karch FE, Lasagna L. Adverse drug reactions: a critical review. JAMA 1975; 234: 1236-41
56. Kane-Gill SL, Kirisci L, Pathak DS. Are the Naranjo criteria reliable and valid for determination of adverse drug reactions in the intensive care unit? Ann Pharmacother 2005; 39: 1823-7
57. Venulet J, Ciucci AG, Berneker GC. Updating of a method for causality assessment of adverse drug reactions. Int J Clin Pharmacol Ther Toxicol 1986; 24: 559-68
58. Benichou C. Criteria of drug-induced liver disorders. Report of an International Consensus Meeting. J Hepatol 1990; 11: 272-6
59. Naranjo CA, Lane D, Ho-Asjoe M, et al. A Bayesian assessment of idiosyncratic adverse reactions to new drugs: Guillain-Barre syndrome and zimeldine. J Clin Pharmacol 1990; 30: 174-80
60. Naranjo CA, Lanctot KL. A consultant's view on the role of Bayesian differential diagnosis in the safety assessment of pharmaceuticals. Drug Inf J 1992; 26: 593-601
61. Naranjo CA, Lanctot KL, Lane DA. The Bayesian differential diagnosis of neutropenia associated with antiarrhythmic agents. J Clin Pharmacol 1990; 30: 1120-7
62. Pere JC, Begaud B, Haramburu F, et al. Computerized comparison of six adverse drug reaction assessment procedures. Clin Pharmacol Ther 1986; 40: 451-61
63. Ennis M, Ohmann C, Lorenz W, et al. Prediction of risk for pseudoallergic reactions and histamine release in patients un-

dergoing anaesthesia and surgery: a computer-aided model using independence-Bayes. Agents Actions 1988; 23: 366-9
64. Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. Eur J Clin Pharmacol 1977; 11: 75-8
65. Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs: II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. J Clin Epidemiol 1993; 46: 1331-6
66. Michel DJ, Knodel LC. Comparison of three algorithms used to evaluate adverse drug reactions. Am J Hosp Pharm 1986; 43: 1709-14
67. Busto U, Naranjo CA, Sellers EM. Comparison of two recently published algorithms for assessing the probability of adverse drug reactions. Br J Clin Pharmacol 1982; 13: 223-7
68. Kramer MS. Assessing causality of adverse drug reactions: global introspection and its limitations. Drug Inf J 1986; 20: 433-7
69. Hutchinson TA, Flegel KM, HoPingKong H, et al. Reasons for disagreement in the standardized assessment of suspected adverse drug reactions. Clin Pharmacol Ther 1983; 34: 421-6
70. Stephens MD. The diagnosis of adverse medical events associated with drug treatment. Adverse Drug React Acute Poisoning Rev 1987; 6: 1-35
71. Naranjo CA, Kwok MC, Lanctot KL, et al. Enhanced differential diagnosis of anticonvulsant hypersensitivity reactions by an integrated Bayesian and biochemical approach. Clin Pharmacol Ther 1994; 56: 564-75
72. Aronson JK, Ferner RE. Clarification of terminology in drug safety. Drug Saf 2005; 28: 851-70
73. Castle W. Adverse drug reactions: scope and limitations of causality assessment and the use of algorithms. Int J Risk Saf Med 1991; 2: 185-91
74. Kelly WN. The quality of published adverse drug event reports: review. Ann Pharmacother 2003; 37: 1774-8

Correspondence: *Taofikat B. Agbabiaka*, Complementary Medicine, Peninsula Medical School, Universities of Exeter and Plymouth, 25 Victoria Park Road, Exeter, EX2 4NT, UK.
E-mail: taofikat.agbabiaka@pms.ac.uk