Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions

Yannick Arimone, Ghada Miremont-Salamé, Françoise Haramburu, Mathieu Molimard, Nicholas Moore, Annie Fourrier-Réglat & Bernard Bégaud

From the INSERM U 657, Département de Pharmacologie, Université Victor Segalen, Bordeaux 2, France

What is already known about this subject

- In pharmacovigilance, many methods have been proposed for causality assessment of adverse drug reactions.
- Expert judgement is commonly used to evaluate the causal relationship between a drug treatment and the occurrence of an adverse event. This form of judgement relies either explicitly or implicitly on causality criteria.

What this study adds

- Our study compares the judgements of five senior experts using global introspection about drug causation and seven causality criteria on a random set of putative adverse drug reactions.
- Even if previous publications have shown poor agreement between experts using global introspection, few have compared judgements of well trained pharmacologists, familiar with using a standardized causality assessment method.

Correspondence
Yannick Arimone, INSERM U 657, Département de Pharmacologie, Université Victor Segalen, 146 Rue Léo Saignat, 33076 Bordeaux cedex, France.
Tel.: + 33 5 5757 1561
Fax: + 33 5 5757 4660
E-mail:
yannick.arimone@etud.u-bordeaux2.fr

Aims
To evaluate agreement between five senior experts when assessing seven causality criteria and the probability of drug causation.

Methods
A sample of 31 adverse event-drug pairs was constituted. For each pair, five experts separately assessed (i) the probability of drug causation, which was secondarily divided into seven causality levels: ruled out (0–0.05), unlikely (0.06–0.25), doubtful (0.26–0.45), indeterminate (0.46–0.55), plausible (0.56–0.75), likely (0.76–0.95), and certain (0.96–1); and (ii) seven causality criteria. To test discrepancies between experts, the kappa index was used.

Results
The agreement of the five experts was very poor (kappa = 0.05) for the probability of drug causation. Among the seven levels of causality, only 'doubtful' showed a significant rate of agreement (kappa = 0.32, $P < 0.001$). For all criteria, the kappa index was significant except for the item 'risk(s) factor(s)' (kappa = 0.09). Agreement between experts was good (0.64, $P < 0.001$) only for the criterion 'reaction at site of application or toxic plasma concentration of the drug or validated test'. However, the rate of agreement with kappa indices of the causality criteria ranged from 0.12 to 0.38.

Conclusions
This study confirms that in the absence of an operational procedure, agreement between experts is low. This should be considered when designing a causality assessment method. In particular, criteria inducing a low level of agreement should have their weight reduced.

Introduction

Many causality methods have been proposed to assess the individual causality between a drug treatment and the occurrence of an adverse event. Roughly, these methods may be classified into three approaches [1, 2]: expert judgement, probabilistic methods and algorithms.

In expert judgement or global introspection, an expert expresses a judgement about possible drug causation after having taken into account all the available and relevant information on the considered case. This approach suffers from marked subjectivity leading to poor reproducibility and intra- and inter-rater disagreements [3–7].

Probabilistic methods are usually regarded as the most rigorous [8]. The probabilistic approach is based on the Bayes theorem and makes it possible to assess the probability (or odds) of drug-causation directly. However, these methods are rather troublesome to use routinely because information for assessing the probability (or odds) of drug causation is rarely available.

Unlike the Bayesian approach, algorithms have appealing simplicity and are much more widely used for the operational assessment of adverse drug reactions (ADRs). The main reason for their use is to increase inter- and intrarater agreement [9, 10].

In these three approaches, several causality criteria are implicitly or explicitly used.

The purpose of this paper was to evaluate the agreement between five senior experts about the assessment of seven criteria, and the global causality assessment of ADRs.

Methods

Adverse drug reactions (ADRs)

The study was conducted on a random sample of 30 ADR cases (31 drug-event pairs): 15 were randomly selected among those collected during a nationwide study and 15 were selected from those received 6 years before (1998) by the Bordeaux Pharmacovigilance Centre [11].

The information for these 30 ADRs was summarized on a standardized form including the characteristics of the patient, the suspected drug with the dates of treatment, the adverse drug effect with the date of onset, major biological and clinical data, the other current medical treatment and the dechallenge.

The experts

Five experts were chosen among of heads of French Regional Centres of Pharmacovigilance or departments of pharmacovigilance of the pharmaceutical industries on the basis of their experience in the field, i.e. at least 10 years of daily routine practice.

The questionnaire

The 30 summaries with corresponding questionnaires were sent to five senior experts.

For each suspected drug in each drug-event pair, the questionnaire asked each expert to assess the criteria as follows:

Time to onset: incompatible/not suggestive/unknown or not available/compatible/highly suggestive.

Dechallenge: against the role of the drug/inconclusive or not available/suggestive.

Rechallenge: negative/not available or inconclusive/positive.

Search for non-drug-related causes: non-drug cause highly probable/not investigated or [and] possible nondrug cause/nondrug cause ruled out.

Risk factor(s) for drug reaction: ruled out or absent/well validated and present.

Reaction at site of application, or toxic plasma concentration of the drug, or validated laboratory test: unrelated or not available/present and/or positive.

Previous information on the drug and symptomatology: type B reaction not previously reported/not available/labelled reaction and/or type A reaction.

Global probability of drug causation was expressed on a 100 mm visual analogue scale (VAS), without graduations from excluded on the extreme left ($P = 0$) to certain on the extreme right ($P = 1$). Expert opinion, expressed by a stroke on the VAS and measured to the nearest millimetre, was considered as the probability of drug causation. In a second step, these probabilities were arbitrarily divided into seven causality levels to make statistical analysis easier: ruled out (0–0.05), unlikely (0.06–0.25), doubtful (0.26–0.45), indeterminate (0.46–0.55), plausible (0.56–0.75), likely [0.76–0.95], and certain [0.96–1].

Statistical analysis

Both for the criteria and causality, the assessment of coding divergences in the case of multiple judges applying a qualitative scale was explored by using the Fleiss kappa index of reliability for multiple categories and multiple experts.

Results

Distribution of adverse events

Many adverse events corresponded to cutaneous effects ($n = 9$). The proportions of the other adverse events were as follows: bleeding ($n = 6$), digestive disorders ($n = 3$),

circulatory system ($n = 3$), respiratory events, fever, hepatic effect, metabolism disorder, anaphylactic shock, cardiac effect, psychic disorder and thrombosis.

Global causality assessment
Table 1 shows the distribution of the 155 causality judgements, e.g. 31 drug-event pairs assessed by five experts according to seven closed causality levels.

For the five experts, the probability distribution shows that clear-cut judgements, i.e. 'ruled out' and 'certain', i.e. probabilities <0.05 or >0.95 were very scarce (Table 1). Indeed, these levels represented 7.7% and 3.2% of 155 judgements. About one third (36.1%), 22.6% to 48.4% according to experts, fell in the 'likely' category. The frequency of this causality level varied between the five experts. Ninety-one out of the 155 assessments (58.7%) were in the 'plausible' or 'likely' categories.

The marked divergences shown in Table 1 were confirmed by the analysis in Table 2.

For the same drug-effect pair, the difference between the two extreme scores among the five assessments ranged from 0.135 to 1 for probability of drug causation and from 0 (one time) to six levels (six times) for the seven predetermined causality levels.

The kappa index for overall results concerning the global probability of drug causation and the analysis per level of causality (Table 3) confirmed that the overall agreement was very poor (i.e. kappa = 0.05), except for the 'doubtful' level, for which the agreement was significant (kappa = 0.32, $P < 0.001$).

Analysis of divergences between experts
As expected, the pattern of assessment was extremely different for the seven causality criteria (Table 4). Some criteria were always informative (e.g. time to onset) and others (e.g. rechallenge) were practically always assessed in the neutral position. Some comments may be made:

The criterion time to onset was mainly assessed as 'positive' although the quotation 'highly suggestive' was rarely found;

The criterion rechallenge seems exceptionally assessable since in 89.3% of cases it was judged 'not available or inconclusive';

The criterion 'search for drug-related causes' does not appear to be informative (80.6% of cases were judged 'not investigated and/or possible nondrug cause');

The criterion 'reaction at site of application or toxic plasma drug concentration or validated laboratory test' was mainly ranked as 'unrelated or not available' *vs.* 'present and/or positive' in 15.3% of cases;

On the other hand, the criteria risk factor(s) for drug reaction and previous information on the drug and symptomatology were generally in favour of the implication of the suspected drug, with 60.3% of quotations for 'well validated and present' and 70.2% for 'labelled reaction and/or type A reaction'.

The same assessment was never or rarely chosen by all five experts on any of the criteria.

The five experts agreed on the same assessment pattern for only three criteria: time to onset, rechallenge, reaction at site of application, or toxic plasma concentration of the drug, or validated laboratory test. Expert 5 tended to give more 'suggestive' quotations for dechallenge than the others, when the others mainly chose 'inconclusive or not available'. Assessments of 'risk(s) factor(s) for the drug causation' were contradictory for the five experts. Experts 3, 4 and 5 had a superimposable distribution for the two possible answers. The assessments of experts 1 and 2 were 'well validated and present' in 90.3% and 74.2% cases, respectively. The assessments were more homogeneous for the criterion dechallenge, except for expert 4, who rarely used the quotation 'against the role of drug'. The neutral quotation 'inconclusive or not available' was found in only 13.3% of cases for expert 5, while it was used in 48.4%, 54.8%, 64.5% and 46.3% of cases for the

Table 1
Distribution of assessments by each expert according to the seven causality levels

	Ruled out (%)	Unlikely (%)	Doubtful (%)	Indeterminate (%)	Plausible (%)	Likely (%)	Certain (%)	Total assessment
Expert1	2 (3.23)	2 (6.67)	0 (0)	3 (9.68)	8 (25.81)	12 (38.71)	4 (12.90)	31 (100)
Expert2	1 (3.23)	1 (3.23)	2 (6.45)	7 (22.58)	5 (16.13)	15 (48.39)	0	31 (100)
Expert3	1 (3.23)	2 (6.67)	4 (12.90)	5 (16.13)	9 (29.03)	9 (29.03)	1 (3.23)	31 (100)
Expert4	6 (19.35)	4 (12.90)	2 (6.45)	4 (12.90)	2 (6.45)	13 (41.94)	0	31 (100)
Expert5	2 (6.45)	5 (16.13)	4 (12.90)	2 (6.45)	11 (35.48)	7 (22.58)	0	31 (100)
Total	12 (7.74)	14 (9.03)	12 (7.74)	21 (38.18)	35 (22.58)	56 (36.13)	5 (3.23)	155 (100)

Table 2

Causality assessment for the 31 adverse drug-event pairs using the VAS and the seven causality levels

| Drug-event pairs | Experts | | | | |
	1	2	3	4	5
1	0.67 (P)	0.89 (L)	0.87 (L)	0.50 (I)	0.74 (D)
2	0.65 (P)	0.59 (P)	0.80 (L)	0.51 (I)	0.22 (U)
3	0.60 (U)	0.89 (L)	0.62 (P)	0.00 (R)	0.49 (I)
4	0.92 (L)	0.92 (L)	0.95 (L)	0.19 (U)	0.74 (D)
5	0.05 (R)	0.15 (U)	0.47 (I)	0.44 (D)	0.60 (D)
6	0.84 (L)	0.54 (I)	0.47 (I)	0.86 (L)	0.22 (U)
7	0.92 (L)	0.91 (L)	0.69 (P)	0.92 (L)	0.70 (D)
8	0.77 (L)	0.94 (L)	0.52 (I)	0.17 (U)	0.70 (D)
9	0.52 (I)	0.52 (I)	0.51 (I)	0.89 (L)	0.69 (D)
10	0.80 (L)	0.83 (L)	0.84.5 (L)	0.92.5 (L)	0.86 (L)
11	0.98 (C)	0.72 (P)	0.00 (R)	0.91 (L)	0.79 (L)
12	0.97 (C)	0.90 (L)	1 (C)	0 (R)	0.85 (L)
13	1 (C)	0.92 (L)	0.80 (L)	0.86 (L)	0.89 (L)
14	0.51 (I)	0.47 (I)	0.63 (P)	0 (R)	0.4 (R)
15	0.77 (L)	0.62 (P)	0.90 (L)	0.54 (I)	0.10 (U)
16	0.68 (P)	0.86 (L)	0.64 (P)	0.05 (R)	0.80 (L)
17	0.75 (P)	0.30 (D)	0.37 (D)	0.86 (L)	0.36 (D)
18	0.78 (L)	0.60 (P)	0.69 (P)	0.82 (L)	0.45 (D)
19	0.48 (I)	0.00 (R)	0.24 (U)	0.20 (U)	0.45 (D)
20	0.55 (P)	0.33 (D)	0.30 (D)	0.30 (D)	0.28 (D)
21	0.97 (C)	0.94 (L)	0.74 (P)	0.87 (L)	0 (R)
22	0.68 (P)	0.54 (I)	0.76 (L)	0.80 (L)	0.65 (D)
23	0.70 (P)	0.54 (I)	0.76 (L)	0.80 (L)	0.65 (D)
24	0.16 (U)	0.92 (L)	0.28 (D)	0.68 (P)	0.68 (D)
25	0.91 (L)	0.95 (L)	0.28 (D)	0 (R)	0.8 (U)
26	0.78 (L)	0.95 (L)	0.63 (P)	0.81 (L)	0.86 (L)
27	0.91 (L)	0.67 (P)	0.87 (L)	0.72 (P)	0.93 (L)
28	0.65 (P)	0.73 (L)	0.61 (P)	0.04 (R)	0.5 (D)
29	0.95 (L)	0.50 (I)	0.86 (I)	0.49 (I)	0.67 (D)
30	0.78 (L)	0.51 (I)	0.12 (U)	0.10 (U)	0.17 (U)
31	0.74 (P)	0.89 (L)	0.66 (P)	0.83 (L)	0.54 (I)

(R): ruled out; (U) unlikely; (D) doubtful; (I) indeterminate; (P) plausible; (L) likely; (C) certain.

Table 3

Kappa index for assessment per causality level

	Overall value	Ruled out	Unlikely	Doubtful	Indeterminate	Plausible	Likely	Certain
Kappa	0.048	−0.039	0.058	0.323	0.064	−0.052	0.049	0.070
P	0.04*	NS	NS	<0.001	NS	NS	NS	NS

other experts. The criterion search for non drug-related causes was mainly assessed as 'not investigated and/or possible non-drug cause' (in 83.3% of cases), except for expert 4 (50%).

As noted above, most assessments for the criterion previous information on the drug and symptomatology were 'labelled reaction and/or type A reaction' except for expert 4 (53.5%).

Criteria	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Time to onset (% (n))	100 (31)	100 (31)	100 (31)	100 (28)	100 (30)
Incompatible (%)	0.00	0.0	0.0	7.1	0.0
Not suggestive (%)	6.5	0.0	6.5	10.7	3.3
Unknown or not available (%)	0.0	9.7	29.0	17.9	13.3
Compatible (%)	74.2	67.7	45.2	5.0	53.3
Highly suggestive (%)	19.4	22.6	19.4	14.3	30.0
Dechallenge (% (n))	100 (31)	100 (31)	100 (31)	100 (28)	100 (30)
Against the role of the drug (%)	0.0	0.0	0.0	21.4	6.7
Inconclusive or not available (%)	48.4	54.8	64.5	46.4	13.3
Suggestive (%)	51.6	45.2	35.5	32.1	80.0
Rechallenge (% (n))	100 (31)	100 (31)	100 (31)	100 (28)	100 (30)
Negative (%)	3.2	0.0	0.0	0.0	10.3
Not available or Inconclusive (%)	87.1	100	87.1	96.4	75.9
Positive (%)	9.7	0.0	12.9	3.6	13.8
Search for non-drug-related causes (% (n))	100 (31)	100 (31)	100 (30)	100 (28)	100(30)
Non-drug cause highly probable (%)	3.2	0.0	3.3	14.3	10.0
Not investigated and/or possible nondrug cause (%)	87.1	93.6	83.3	50.0	86.7
Non-drug causes ruled out (%)	9.7	6.4	13.3	35.7	3.3
Risk factor(s) for drug reaction (% (n))	100(31)	100(31)	100(31)	100(28)	100 (30)
Ruled out or absent (%)	9.7	25.8	58.1	50.0	56.7
Well validated and present (%)	90.3	74.2	41.9	50.0	43.3
Reaction at site of application, or plasma concentration of the drug known as toxic, or validated laboratory test (% (n))	100 (31)	100 (31)	100 (31)	100 (28)	100 (30)
Unrelated or not available (%)	83.9	80.7	83.9	85.7	90.0
Present and/or positive (%)	16.1	19.4	16.1	14.3	10.0
Previous information on the drug and symptomatology (% (n))	100 (31)	100 (31)	100 (31)	100 (28)	100 (30)
Reaction not previously reported and type B reaction (%)	6.5	6.5	22.6	0.0	16.7
Not available (%)	25.8	12.9	9.7	46.4	3.3
Labeled reaction and/or type A reaction (%)	67.7	80.7	67.7	53.6	80

Table 4

Assessments of causality criteria for the five experts

The rate of agreement within the group of experts varied according to the item under evaluation. In spite of this variation, inter-rater agreement remained low (Table 5). Indeed, except for the criterion reaction at site of application, or toxic plasma concentration of the drug, or validated laboratory test (kappa = 0.642), the kappa index of the six other causality criteria ranged from 0.09 to 0.38.

The detailed analysis of criteria using more than two quotations (e.g. time to onset) showed that the highest rate of agreement was obtained by 'the highly suggestive' time to onset and 'not available or not conclusive' and by the 'positive' rechallenge (respectively 0.53, 0.40 and 0.45). The weakest agreements were found for the items risk factors of drug causation and search for non-drug-related causes (respectively 0.08, 0.12).

Table 5

Kappa index for seven causality criteria

Criteria	Kappa	P
Time to onset	0.264	<0.001
Incompatible	–0.020	NS
Not suggestive	0.020	NS
Unknown or not available	0.168	<0.01
Compatible	0.193	<0.001
Highly suggestive	0.531	<0.001
Dechallenge	0.226	<0.001
Against the role of the drug	0.003	NS
Inconclusive or not available	0.266	<0.001
Suggestive	0.231	<0.001
Rechallenge	0.379	<0.001
Negative	0.093	NS
Not available or Inconclusive	0.400	<0.001
Positive	0.452	<0.001
Search for non drug-related causes	0.124	<0.01
Non-drug cause highly probable	0.173	<0.01
Not investigated and/or possible nondrug cause	0.089	NS
Non-drug causes ruled out	0.146	<0.01
Risk factor(s) for drug reaction	0.088	NS
Reaction at site of application, or plasma concentration of the drug known as toxic, or validated laboratory test	0.642	<0.001
Previous information on the drug and symptomatology	0.258	<0.001
Reaction not previously reported and type B reaction	0.191	<0.001
Not available	0.127	<0.05
Labelled reaction and/or type A reaction	0.385	<0.001

Discussion

Three approaches are mainly used to assess the causal relationship between drug treatment and the occurrence of adverse events: (i) expert judgement (global introspection), (ii) probabilistic approaches and (iii) algorithms. All use implicit or explicit analysis of causality criteria. Our aim was to measure the agreement between five experts when assessing seven causality criteria and the probability of drug causation. The probability pattern of drug causation for all five experts showed that drugs were mainly considered as a plausible or likely cause of the adverse reaction. This result was expected since cases are reported to pharmacovigilance centres only when the physician already suspects drug causation for the adverse event. On average, agreement about the probability of drug causation was very poor, a good rate of agreement being found only for the 'doubtful' level.

Several factors could account for the observed distribution of assessments:

For the criterion time to onset, the proportion of 'compatible' assessments was an expected fairly high, since a compatible chronology is a core prerequisite for suspecting causality in pharmacovigilance, i.e. drug precedes the reaction. A 'highly suggestive' onset corresponds to a reaction occurring during drug administration (e.g. anaphylactic shock), which is a much rarer scenario.

The criterion rechallenge was only rarely assessable because it was only rarely attempted.

The criterion search for non-drug related causes was not informative as the search was never judged complete enough to rule out another possible explanation for the adverse event.

As expected, the criteria reaction at site of application, or toxic plasma concentration of the drug or validated laboratory test gave a high rate of agreement but were rarely found in the reported cases.

In most cases, the criteria risk factor and previous information on the drug and symptomatology were considered to plead in favour of the implication of the drug.

As stated above, the level of agreement found in this study was relatively low (kappa index ranging from 0.09 to 0.38), except for two criteria (rechallenge and reaction at site of application, or toxic plasma concentration of the drug, or validated laboratory test. The criteria risk factor(s) and search for non-drug-related causes explained most of the disagreement between experts: as opposed to factual criteria (rechallenge and reaction at site of application, or toxic plasma concentration of the drug, or validated laboratory test) [12–14], they call upon introspection and result in more divergence of opinion [15]. The factual criteria are the evaluation of presence or not of a symptom of an action of treatment whereas the other causality assessment criteria can be regarded as already an objective explanation of the biological process leading to the occurrence of an adverse event. The typical cases are the criteria 'risk factors' or 'search for nondrug related causes'. For these two criteria, the evaluation is made upon an objective assumption of relationship adverse event/drug, in which a more or less clarified explanatory mechanism, including the drug or not, leads to the occurrence of adverse event. In this case, the evaluation of criteria belongs to a search of confirmation or invalidation of the mechanism. The difference between these two types of introspection could explain the levels of agreement found in the results.

Divergences between experts assessing ADRs have already been found in other studies. Karch *et al.*

[5] found that three clinical pharmacologists agreed on only 50% of cases. In the study of Koch-Weser *et al.* [4], agreement was good for the drugs likely to have caused the ADR. In both of those studies, the main reasons for disagreement were related to the information pattern of the suspected ADR. Indeed, the absence of a discriminating criterion makes clear-cut opinion difficult to reach (i.e. extreme causality: 'ruled out' or 'certain') [12, 16–19]. This purpose pleads for the use of a combination of the causal criteria in a standardized procedure, like algorithms, which improves the reliability of global causality assessment [1–3, 9]. For the assessment of serious cases, in a decision making context, more reliability approaches should be used such as step by step consensual approach [3].

In conclusion, this study confirms that disagreement exists between experts in the assessment of global causality or causality criteria [20], mainly owing to difficulty in assessing criteria that are not solely based on facts [21]. These findings should be considered when designing new assessment procedures. In particular, the weight of criteria associated with a low rate of agreement should be reduced.

References

1 Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. Drug Saf 1997; 17: 374–89.

2 Stephens MD. The diagnosis of adverse medical events associated with drug treatment. Adverse Drug React Acute Poisoning Rev 1987; 6: 1–35.

3 Kramer MS. Difficulties in assessing the adverse effects of drugs. Br J Clin Pharmacol 1981; 11(Suppl 1): 105S–110S.

4 Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. Eur J Clin Pharmacol 1977; 11: 75–8.

5 Karch FE, Smith CL, Kerzner B, Mazzullo JM, Weintraub M, Lasagna L. Adverse drug reactions – a matter of opinion. Clin Pharmacol Ther 1976; 19: 489–92.

6 Feinstein AR. Clinical biostatistics. XXX. Biostatistical problems in 'compliance bias'. Clin Pharmacol Ther 1974; 16: 846–57.

7 Blanc S, Leuenberger P, Berger JP, Brooke EM, Schelling JL. Judgments of trained observers on adverse drug reactions. Clin Pharmacol Ther 1979; 25: 493–8.

8 Lanctot KL, Naranjo CA. Comparison of the Bayesian approach and a simple algorithm for assessment of adverse drug events. Clin Pharmacol Ther 1995; 58: 692–8.

9 Hutchinson TA, Lane DA. Assessing methods for causality assessment of suspected adverse drug reactions. J Clin Epidemiol 1989; 42: 5–16.

10 Begaud B. Dictionary of pharmacoepidemiology. Chichester: Wiley, 2000.

11 Pouyanne P, Haramburu F, Imbs JL, Begaud B. Admissions to hospital caused by adverse drug reactions: cross sectional incidence study. French Pharmacovigilance Centres. BMJ 2000; 320: 1036.

12 Pere JC, Begaud B, Haramburu F, Albin H. Computerized comparison of six adverse drug reaction assessment procedures. Clin Pharmacol Ther 1986; 40: 451–61.

13 Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, Janecek E, Domecq C, Greenblatt DJ. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981; 30: 239–45.

14 Macedo AF, Marques FB, Ribeiro CF, Teixeira F. 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 Safety 2005; 14: 885–90.

15 Hutchinson TA, Flegel KM, HoPingKong H, Bloom WS, Kramer MS, Trummer EG. Reasons for disagreement in the standardized assessment of suspected adverse drug reactions. Clin Pharmacol Ther 1983; 34: 421–6.

16 Miremont G, Haramburu F, Begaud B, Pere JC, Dangoumau J. Adverse drug reactions: physicians' opinions versus a causality assessment method. Eur J Clin Pharmacol 1994; 46: 285–9.

17 Arimone Y, Begaud B, Miremont-Salame G, Fourrier-Reglat A, Moore N, Molimard M, Haramburu F. Agreement of expert judgment in causality assessment of adverse drug reactions. Eur J Clin Pharmacol 2005; 61: 169–73.

18 Venulet J, Ciucci A, Berneker GC. Standardized assessment of drug–adverse reaction associations – rationale and experience. Int J Clin Pharmacol Ther Toxicol 1980; 18: 381–8.

19 Meyboom RH. Causality assessment revisited. Pharmacoepidemiol Drug Safety 1998; 7(Suppl 1): S63–5.

20 Leventhal JM, Hutchinson TA, Kramer MS, Feinstein AR. An algorithm for the operational assessment of adverse drug reactions. III. Results of tests among clinicians. JAMA 1979; 242: 1991–4.

21 Macedo AF, Marques FB, Ribeiro CF, Teixeira F. 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.