PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2005; **14**: 885–890
Published online 29 July 2005 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/pds.1138

ORIGINAL REPORT

Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel†

Ana Filipa Macedo*, Francisco Batel Marques, Carlos Fontes Ribeiro and Frederico Teixeira

Núcleo de Farmacovigilância do Centro, Faculdade de Medicina, Faculdade de Farmácia, Universidade de Coimbra, Administração Regional de Saúde do Centro, Portugal

SUMMARY

Purpose To compare the results of causality assessments of reported adverse drug reactions (ADR's) obtained from decisional algorithms with those obtained from an expert panel using the WHO global introspection method (GI) and to further evaluate the influence of confounding variables on algorithms ability in assessing causality.

Method Two hundred sequentially reported ADR's were included in this study. An independent researcher used algorithms, while an expert panel assessed the same reports using the GI, both aimed at evaluating causality. Reports were divided into three groups according to the presence, absence or lack of information on confounding variables.

Results For the total sample, observed agreements between decisional algorithms compared with GI varied from 21% to 56%, average of 47%. When confounding variables were taken into account, agreements varied between 41% and 69%, average of 58%; 8% and 65%, average of 46% and 15% and 53%, average of 42% accordingly to the absence, lack of information or presence of confounding variables, respectively. The extend of reproducibility beyond chance was low for the total sample (average Kappa = 0.26) and within the groups considered.

Conclusion The overall observed agreement between algorithm and GI was moderate although poorly different from chance, confounding variables being a shortcoming of algorithms ability in assessing causality. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS — pharmacovigilance; adverse drug reaction; causality assessment; algorithm; global introspection

INTRODUCTION

An adverse drug reaction is any noxious and unintended response to a drug that occurs at doses normally used in man for prophylaxis, diagnosis, therapy or modification of a physiological function.[1] Causality assessment of identified adverse events during drug exposure is crucial in pharmacovigiliance activities due to its implications on the risk-benefit ratio evaluations of medicines.

The 'Núcleo de Farmacovigilncia do Centro', NFC—the central Portugal regional pharmacovigilance unit—started the reception of ADR reports from family physicians and community pharmacists in January 2001. According to its regulatory responsibilities, causality assessment of reported ADR's is mandatory. As a component of the adopted standard operating procedures of the Portuguese pharmacovigilance system, the global introspection (GI) based on the World Health Organisation scale of imputability[2] has been used.

* Correspondence to: Ana Filipa Macedo, Faculdade de Ciências da Saúde. Universidade da Beira Interior, 6201-001 COVILHÃ. Portugal. E-mail: filipa@fcsaude.ubi.pt

†Ana Filipa Macedo is PhD student supported by a research grant from ''Fundação para a Ciência e a Tecnologia—programa Praxys BD/758/2000—Portugal''.

Received 8 July 2002
Revised 9 June 2005
Accepted 14 June 2005

Despite its usefulness, the GI method has been subject to criticisms of subjectivity and imprecision since it is mainly based on expert clinical judgements.[3,4]

Concurrently, but not alternatively, several decisional algorithms have been published, combining, and scoring different criteria as an explicit approach, claiming the advantage of avoiding subjective bias.[3–5] However, none of the algorithms published since 1976 have been universally accepted as a gold standard and several studies pointed out disagreements between the results obtained from the use of the different algorithms in assessing causality for the same ADR reports.[6–12]

However, it should be emphasised that algorithms were not designed and do not intent to replace medical diagnosis.

This study was carried out to evaluate causality in reported adverse drug reactions (ADR) using published algorithms and to further compare the results with those obtained by the consensus opinion of specialist's panel using the World Health Organisation GI method in order to determine the usefulness of algorithms on causality assessment of ADR.

METHODS

Two hundred sequentially reported ADR's to the NFC, 131 from general practitioners (GPs) and 69 from community pharmacists, from January 2001, were studied. The sample included a large spectrum of clinical events. For each ADR report, an average of two clinical manifestations was described. Twenty-seven percent of the reported ADR's were considered serious according to the WHO's criteria.[13] One hundred and forty eight different drugs were reported as responsible for the suspected ADR's. In 78% of the reports, concurrent medications were present (maximum 7, average 2).

From a literature search on Pubmed using 'pharmacovigilance', 'imputability', 'algorithm', 'introspection' and 'adverse reaction' mesh headings, the following decisional algorithms published between 1970 and 2001 were identified:

AD, ADRIAN[14]
Aust, Australian[15]
By, Bayesian[16]
B, Blanc[4]
Ca, Castle[17]
Co, Cornelli[18]
CPMP, ABO system[19]
D, Dangoumau[20,21]
Em, Emanueli[22]
Ev, Evreux[23]
HM, Hoskins & Mannino[24]
HS, Hsu-Stoll[25]
I, Irey[26]
Ja, Jain[27]
Jo, Jones[28]
KL, Karch & Lasagna[5]
Ki, Kitaguchi[29]
Kr, Kramer[30–32]
La, Lagier[33]
Lu, Loupi[34]
MV, Maria V.[35]
N, Naranjo[36]
R, RUCAM[37]
St, Stephens[38]
Sk, Stricker[39]
T, Taiwan[40]
V, Venulet[41]
W, Weber[42]
WHO[2]

Some algorithms were withdrawn from the study. Adrian, Castle and Evreux algorithms, which present results in numerical values were excluded due to the lack of equivalence with causality terminology used in the GI method. Bayesian, Lagier and Hoskins & Mannino methods need aetiologic balance in causality assessment, information not requested in filling Portuguese ADR report forms. Algorithms aimed at specific ADR's were also withdrawn from the study—hepatic toxicity (Maria V. and Stricker), teratogenicity (Loupi) and predetermined disease states (RUCAM). Finally, we excluded Jain, Taiwan and CPMP approaches, which only have three levels for causality assessment, therefore giving inconsistency in matching results with the six levels of the GI method.

The following algorithms were then applied, all supported by the combination of five common major criteria for causality assessment: challenge, dechallenge, rechallenge, previous bibliographic description and etiologic alternatives.

Aust, Australian[15]
B, Blanc[4]
Co, Cornelli[18]
D, Dangoumau [20,21]
Em, Emanueli[22]
HS, Hsu Stoll[25]
I, Irey[26]
Jo, Jones[28]
KL, Karch & Lasagna[5]
Ki, Kitaguchi[29]
Kr, Kramer[30–32]
N, Naranjo[36]
St, Stephens[38]
V, Venulet[41]
W, Weber[42]

An investigator independently assessed the causality of the 200 adverse events using the decisional algorithms. ADR's reports were simultaneously and independently assessed by an expert panel comprising two clinical pharmacologists (one also being a specialist in internal medicine), two pharmacists and a general practitioner using the GI method based on the World Health Organisation scale of imputability.[2]

Causality assessments produced from both modalities were finally compared and analysed using the kappa index of reliability. Epi Info 2002 software was used for the analysis.

The terminology used to express causality was found to be qualitatively and quantitatively different between algorithms, therefore giving limitations to accurate comparisons. In the light of such findings, a correspondence table of terms from different algorithms was developed (Table 1). Other correspondences between terms were published,[43,44] but so far they did not include all the studied algorithms. According to the GI method, there are six levels of imputation: certain/definitive, probable, possible, unlikely, conditional and unclassifiable. A drug-related event can only be considered when one at the first three levels is found. The equivalence of terms between the selected algorithms and GI was based on the following criteria:

Table 1. Correspondence between causality assessment terms

Co	Hs	WHO	I	KL	D	B	Ki	Em	V	Kr	N	Jo	Ast	W	St
Unrelated	Unrelated	Unclassifiable	Negative	Unrelated	Excluded	Doubtful	Unknown	Unrelated	Unrelated	Unlikely	Doubtful	Remote	General List	Unlikely	Doubtful
Very Doubtful	Remote	Conditional													
Possible	Possible	Unlikely	Coincidental	Conditional	Doubtful	Coincidental	Remote	Possible	Unlikely						
Probable	Probable	Possible	Possible	Possible	Possible	Possible	Possible	Probable	Possible	Possible	Possible	Possible	Possible	Possible	Possible
Almost Definite	High Probable	Probable	Probable	Probable	Likely	Probable	Probable	Almost Definite	Probable	Probable	Probable	Probable	Probable	Probable	Probable
Definite Risk	Definite	Certain	Causative	Definite	Very Likely	Certain	Definite	Definite	Definite	Definite	Definite	High Probable	Certain	Definite	Almost Certain

When, for a given algorithm, a level of imputation between 'certain' and 'probable' was present, causality was considered as 'probable'. The level of causality 'possible' was considered the lower limit of acceptance of an adverse event as an adverse drug reaction. In order to classify an adverse event as non-drug related, the following terms were considered as equivalents, since different terminologies are used with the same meaning across the selected algorithms: 'coincidental', 'conditional', 'excluded', 'general list', 'negative', 'remote', 'unclassifiable', 'unknown', 'unlikely', 'unrelated' and 'very doubtful'.

Difficulties in establishing causality assessments with decisional algorithms are often due to the presence of the so-called 'confounding variables'. 'Confounding occurs when the estimate of a measure of association between drug exposure and health status is distorted by the effect of one or several other variables that are also risk factors for the outcome of interest'.[45] Confounding variables included characteristics of adverse events which distorted (strengthen or weaken) the association between a given algorithm imputation and the '*Gold Standard*' (compromising algorithms validity), measured by their agreement, because they made the two methods unadjusted concerning information assessed and its relative ponderation. Confounding variables were considered as: underlying disease, concomitant use of other drugs, absence of bibliographic description, dechallenge with simultaneous treatment, unknown dechallenge, absence of dechallenge, negative dechallenge, dechallenge with simultaneous withdrawn of concomitant drugs, unknown challenge and negative challenge. In order to evaluate the influence of such variables on the concordance between GI method and algorithms, ADR reports were further evaluated and grouped according to the presence of confounding variables. Concordance between algorithms and GI method was also performed within groups.

RESULTS

According to the expert panel, 29 cases were classified as *Definitive,* 77 as *Probable*, 59 as *Possible* and 5 as *Unlikely*; 18 reports were under follow-up evaluation, therefore being classified as *Conditional*, while the presence of insufficient or contradictory information lead to 12 *Unclassifiable* cases. Results of comparisons between causality assessments from the GI method and decisional algorithms are presented on Table 2. For the total sample (200 ADR reports), observed agreements between the GI method and decisional algorithms varied from 21% (Co) to 56%

Table 2. Observed proportion (%) of agreement between Algorithms and Global Introspection (G.I.) and proportion of agreement expected by chance (kappa index)

	I	KL	D	B	Kr	N	Jo	Ki	Ast	W	Em	Co	St	V	HS	Mean
Total ADR's Reports ($n = 200$)	55%, 0.33	41%, 0.23	51%, 0.31	46%, 0.22	54%, 0.35	49%, 0.23	47%, 0.24	56%, 0.36	53%, 0.32	40%, 0.25	54%, 0.32	21%, 0.005^{NS}	41%, 0.19	45%, 0.24	49%, 0.24	47%, 0.26
Group A ($n = 54$)	56%,*	65%, 0.28	61%, 0.30	65%, 0.28	65%, 0.30	52%, 0.002^{NS}	65%, 0.28	69%, 0.38	65%, 0.28	41%, 0.17	67%, 0.33	41%, 0.02^{NS}	43%, 0.00^{NS}	50%, 0.12^{NS}	63%, 0.26	58%, 0.21
Group B ($n = 26$)	62%, 0.34	58%, 0.10^{NS}	38%, 0.06^{NS}	58%, 0.20	58%, 0.39	50%, 0.27	58%, 0.24	54%, 0.25	38%, 0.02^{NS}	38%, 0.12^{NS}	65%, 0.44	8%, -0.06^{NS}	31%, -0.03^{NS}	38%, 0.06^{NS}	46%, 0.07^{NS}	46%, 0.16
Group C ($n = 120$)	53%, 0.27	27%, 0.13	48%, 0.23	34%, 0.14	48%, 0.20	48%, 0.13	37%, 0.14	51%, 0.22	51%, 0.21	40%, 0.18	46%, 0.20	15%, -0.02^{NS}	42%, 0.23	43%, 0.20	43%, 0.15	42%, 0.17

Results express the percent of observed agreement and the coefficients of agreement significantly different from chance ($p < 0.05$) between causality assessments given by different decisional algorithms and those obtained by global introspection, in a total of 200 adverse reports and in three subgroups identified according to confounding variables: Group A—absence of confounding variables; Group B—unknown confounding variables; Group C—presence of confounding variables. NS—not statistically significant.
*Kappa statistics cannot be computed because the algorithm assessed one single category of causality.

(Ki), the average being 47%, although agreements beyond what would be expected by chance varied from very low and not statistically different from chance with Cornelli algorithm ($\kappa = 0.005$), to fair agreement ($\kappa = 0.36$; $p < 0.05$) with Kitaguchi algorithm, the average kappa being 0.26.

When confounding variables were taken into account, three groups were found: group A (ADR's reports without any confounding variable, $n = 54$), group B (ADR's reports were definitive information on confounding variables could not be collected, $n = 26$) and group C (ADR's reports with at least one confounding variable, $n = 120$). For Group A observed agreements varied between 41% (Co and W) and 69% (Ki), the average being 58%, with poor and not statistically different from chance agreement ($\kappa = 0.00$ and $\kappa = 0.02$) with Stephens and Cornelli algorithms and fair agreement ($\kappa = 0.38$; $p < 0.05$) with Kitaguchi algorithm, the average kappa being 0.21. The kappa statistic was not computed for Irey algorithm because this method assessed Group A reports into a single category of causality.

For Groups B and C observed results varied from, respectively, 8% ($\kappa = -0.05$) (Co) to 65% ($\kappa = 0.44$; $p < 0.05$) (Em) with an average of 46% observed agreement and from 15% ($\kappa = -0{,}02$) (Co) to 53% ($\kappa = 0.27$; $p < 0.05$) (I), with an average of 42% observed agreement.

DISCUSSION AND CONCLUSIONS

The GI method was selected as the standard for comparisons in the absence of a well established gold standard. Its internal validity lies on the small probability of different results in the consensus obtained for similar conditions, whilst its external validity remains to be studied.

The extent of agreement between decisional algorithms and GI was analysed using the kappa statistic. This quantifies the extent to which the observed proportion of agreements exceeds the proportion of agreements we would expect by chance alone.[46]

Full agreement with GI was not found for any algorithm. A large range of results was found amongst algorithms providing evidence that they are not interchangeable.

The overall observed agreement between algorithms and GI was moderate (average 47%), although poorly different from chance (average $\kappa = 0.26$). In the absence of confounding variables, agreements in the observed results were found to be high when compared to the total sample but for Dangoumau, Kramer, Naranjo, Australian, Weber, Cornelli, Stephens and

Venulet algorithms such difference on degree of agreement might be due to chance, considering the reduction in kappa values. Confounding variables were found to compromise algorithms ability to establish causality of reported ADR's by decreasing the observed concordance of their results with those obtained from an expert panel using the GI method and increasing the extent of agreement expected as a result of chance.

The algorithms presenting the highest levels of observed agreement were also the one's expressing the highest agreement beyond chance. This was also found when comparisons between groups were analysed. Kitaguchi, Emanueli and Kramer algorithms were found to provide the best global agreements with GI method even in the presence of confounding variables.

Cornelli, Stephens and Weber judges agreed with GI less often than chance would predict. These algorithms do not offer advantages in the causality assessment process.

In the light of the present findings, the value of decisional algorithms in the assessment of ADR's causality remains to be established, confounding variables being a shortcoming of their usefulness.

REFERENCES

1. Adverse Reaction definition. Available at http://www.who-umc.org/defs.html, January 2002.
2. World Health Organization causality assessment method. Available at http://www.who-umc.org/defs.html, January 2002.
3. Karch FE, Smith CL, Kerzner B, *et al.* Adverse drug reactions a matter of opinion. *Clin Pharmacol Ther* 1976; **19**: 489–492.
4. Blanc S, Levenberger P, Berger JP, Brooke EM, Schelling JL. Judgements of trained obsevers on adverse drug reactions. *Clin Pharmacol Ther* 1979; **25**: 493–498.
5. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247–254.
6. Hutchinson TA, Flegel KM, Herbert H, *et al.* Reasons for disagreement in the standardized assessment of suspected adverse drug reactions. *Clin Pharmacol Ther* 1983; **34**: 421–426.
7. Dangoumau J, Begaud B, Boisseau A, Albin H. Methodes d'identification et d'imputabilité des effets indésirables des medicaments. L'effet indésirable, une realité fuyante? *Therapie* 1980; **35**: 287–292.
8. Begaud B, Boisseau A, Albin H, Dangoumau J. Comparaison de quatre méthodes d'imputabilité des effets indésirables des médicaments. *Therapie* 1981; **36**: 65–70.
9. Leroy O, Begaud B, Dangoumau J, Peytour P, Salamon R. Etude comparative de quatre méthodes d'imputabilité. *Therapie* 1981; **36**: 223–227.
10. Pere JC, Begaud B, Haramburu F, Albin H. Computarized comparison of six adverse drug reactions assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–461.
11. Girard M. L'imputation en pharmacovigilance est-elle fiable? *Therapie* 1984; **39**: 291–296.
12. Pere JC, Begaud B, Haramburu F, Albin H. Méthodes d'étude des effets indésirables des médicaments.II. Profil et comparaison de cinq méthodes d'imputabilité. *Therapie* 1984; **39**: 369–378.
13. World Health Organization criteria for serious adverse event or reaction. Available at http://www.who-umc.org/defs.html, January 2002.
14. Castle WM. Adverse drug reactions: scope and limitations of causality assessment and the use of algorithms. *Int J Risk Safety Med* 1991; **2**: 185–192.
15. Mashford ML. The Australian method of drug-event assessment. *Drug Inf J* 1984; **18**: 271–273.
16. Auriche M. Approache Bayésienne de l'imputabilitées phénomenes indésirables aux médicaments. *Therapie* 1985; **40**: 301–306.
17. Castle WM. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 297–302.
18. Stephens MD. The diagnosis of adverse medical events associated with drug treatment. *Adverse Drug React Acute Poisoning Rev* 1987; **1**: 1–35.
19. Meyboom RHB, Royer RJ. Causality classification at Pharmacovigilance Centres in European Community. *Pharmacoepedemiol Drug Safe* 1992; **1**: 87–97.
20. Dangoumau J, Evreux JC, Jouglard J. Méthode d'imputabilité des effets indésirables des médicaments. *Therapie* 1978; **33**: 373–381.
21. Begaud B, Evreux JLC, Jouglard JC, Lagier G. Imputabilité des effets inattendus ou toxiques des médicaments: Actualisation de la méthode utilisée en France. *Therapie* 1985; **40**: 111–114.
22. Emanueli A. A simple algorithm for assessing causality of adverse reactions. Special Workshop-Industrial. *Drug Inf J* 1984; **18**: 303–306.
23. Evreux JC, Loupi E, Descotes J, Ventre JJ. Evaluation des dossiers d'effets indésirables d'origine médicamenteuse. Proposition.2. L'imputabilité. *Therapie* 1982; **37**: 657–670.
24. Hoskins RE, Mannino S. Causality assessment of adverse drug reactions using decision support and information tools. *Pharmacoepidemiol Drug Safe* 1992; **1**: 235–249.
25. Hsu P-H, Stoll RW. Causality assessment of adverse events in clinical trials: II. An algorithm for drug causality assessment. *Drug Inf J* 1993; **27**: 387–394.
26. Irey NS. Tissue reactions to drugs. *Am J Pathol* 1976; **82**: 617–648.
27. Jain KK. A short practical method for triage of adverse drug reactions. *Drug Inf J* 1995; **29**: 339–342.
28. Jones JK. Adverse drug reactions in the community health setting: approaches to recognizing, counselling and reporting. *Fam Community Health* 1982; **5**: 58–67.
29. Kitaguchi T, *et al.* Some Assessment Systems for Industry Post marketing Adverse Drug Reaction (ADR) Information. *Iyakuhin Kenkyu* 1983; **14**: 980–993.
30. 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**: 633–638.
31. Kramer MS, Leventhal JM, Hutchinson TA, Feinstein AR. An algorithm for the operational assessment of adverse drug reactions.I. Background, description and instructions for use. *Jama* 1979; **242**: 623–632.
32. 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**: 1991–1994.

33. Lagier G, Vicens M, Castot A. Imputabilité en pharmacovigilance. Principes de la méthode apreciative pondérée (MAP) et principales erreurs à éviter. *Therapie* 1983; **38**: 303–318.
34. Loupi E, Ponchon AC, Ventre JJ, Evreux JCI. Imputabilité d'un effet tératogène. *Therapie* 1986; **41**: 207–210.
35. Maria V, Victorino R. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* 1997; **26**: 664–669.
36. Naranjo CA, Busto U, Sellers EM, *et al.* A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–245.
37. Bénichou C, Danan G. A new method for drug causality assessment: RUCAM. In *Adverse Drug Reactions: A Practical Guide to Diagnosis and Management*, Christian B (ed.). John Wiley & Sons Ltd.: Chichester, UK, 1994; 277–283.
38. Stephens MD. Assessment of causality in industrial setting. Special Workshop-Industrial. *Drug Inf J* 1984; **18**: 307–313.
39. Stricker BH. Diagnosis and causality assessment of drug-induced hepatic injury. In *Drug Induced Hepatic Injury*, Dukes MNG (ed.). Elsevier: Amesterdam, 1989; 1–13.
40. Taiwan (Triage Application for Imputologists Without An Interesting Name). In *Detection of New Adverse Drug Reactions*, 4th edn, Stephens MDB, Talbot JCC, Routledge PA (eds). Macmillan Reference LTD 2000: Basingstoke, UK; appendix 3.
41. Venulet J, Ciucci A, Berneker GC. Updating of a method for causality assessment of adverse drug reactions. *Int J Clin Pharmacol Ther Toxicol* 1986; **24**: 559–568.
42. Weber JCP. Storage and retrieval of data on adverse reactions to drugs. *Interphex Symposium Brighton* 1980.
43. Lagier G, Vincens M, Lefebure B, Frelon JH. Imputation médicament par médicament en pharmacovigilance. Essai de comparaison des différentes méthodes. *Therapie* 1983; **38**: 295–302.
44. Loupi E, Ponchon AC, Ventre JJ, Descotes J, Evreux JCI. Le rechallenge est-il nécessaire à une imputabilité maximale? Valeur comparée dans sept méthodes d'imputabilité. *Therapie* 1984; **39**: 461–466.
45. Strom BL. Bias and confounding in Pharmacoepidemiology. In *Pharmacoepidemiology*, 3rd edn, Strom BL (ed.). John Wiley Sons, Ltd.: Chichester, UK, 2000.
46. Gordis L. Assessing the validity and reliability of diagnostic and screening tests. In *Epidemiology*, Gordis LWB (ed.). Saunders Company: Elsevier, Philadelphia, USA, 2000.

