Research Letter

Evaluation of inter-rater agreement between three causality assessment methods used in pharmacovigilance

Sir,

Causality assessment is the evaluation of a possibility that a particular treatment is the cause of an adverse event.[1] It plays a key role in pharmacovigilance both toward signal generation and risk benefit evaluation. There are several methods and algorithms for causality assessment, but none is considered the gold standard and use varies from country to country. This is due to inter-individual variation in using the methods as also varying sensitivities and specificities of each method.[2,3] Bayesian methods of causality have greater credence but have lower acceptance as they are rigorous to perform and involve use of complex calculations.[2,4] In the absence of a universally accepted method the present study was carried out to assess the intra- and inter-rater variations among three widely used causality methods—the WHO-UMC method (recommended by the Pharmacovigilance Program of India (PvPI)), Naranjo's algorithm (a widely quoted method in case reports in literature) and the European ABO method (used in European Union to harmonize decision making). To the best of our knowledge, a study comparing these three methods has never been carried out.

The study protocol was deemed exempt from review by the Institutional Ethics Committee. A total of 401 adverse events collected as part of the surveillance of the PvPI between January and April 2013 were assessed independently by three different raters who had at least 3 years of training in the discipline. The WHO- UMC method classifies causality as certain, probable, possible, unlikely, unclassified and unclassifiable, that is, six categories. Naranjo's algorithm categorizes causality into four categories—definite, probable, possible and doubtful. The European ABO system also categorizes causality into three categories—A, B and O.[1-3] Taking into consideration these differences, agreement was assessed by collapsing the categories and making them uniform. While comparing the agreement between European ABO method with the other two methods, all scales were classified into three ranks as follows: (a) certain/probable/A (rank 1); (b) possible/possible/B (rank

2) and (c) unlikely/doubtful/unclassified/unclassifiable/O (rank 3). However, an agreement between WHO-UMC scale and Naranjo's algorithm was made using four ranks—(a) certain/definite (rank 1); (b) probable (rank 2); (c) possible (rank 3) and (d) Unlikely/unclassified/unclassifiable/ doubtful (rank 4). Ranks were compared with each other and agreement was calculated using Fleiss kappa index.[5,6]

A majority of the ADRs found by WHO-UMC method were possible (39.06%), followed by probable (34.33%) and certain (26.18%). Naranjo's algorithm found 52.54% of the ADRs as probable followed by possible (47.30%). Causality assessment by European ABO method assessed 56.87% of the ADRs as A followed by B (42.86%). When the three methods were compared, the highest inter-rater agreement was obtained with the Naranjo's algorithm (314/401 = 78%). The overall inter-rater agreement in all three methods was high (kappa statistic = 0.69) [Table 1]. Overall, intra-rater agreement was substantial in two raters (kappa = 0.69 and 0.74) and moderate (kappa = 0.66) in the third rater. The highest agreement was observed between Naranjo's algorithm and European ABO method (kappa = 0.74) followed by WHO-UMC scale and Naranjo's algorithm (kappa = 0.69) on pair-wise comparisons.

The present study found good inter- and intra-rater concordance between the three commonly used causality methods—the WHO causality scale, the Naranjo's algorithm and the European ABO system. Davies et al.[7] observed a fair agreement (kappa = 0.31) between six assessors using the WHO-UMC method, Naranjo's algorithm and Venulet algorithm which was lower than that observed in the present study. A higher agreement between the WHO-UMC method and Naranjo's algorithm was reported by them which has been corroborated by our observation as well (kappa = 0.61).[7] Arimone et al.[8] found poor agreement between five raters in 31 adverse drug reaction pairs using seven criteria for causality assessment and global causality derived from a questionnaire.[8] Most common causality rating in our study with the WHO-UMC method was "possible" which corroborates with Macedo et al.,[9] while Naranjo's algorithm assessed most common causality

Table 1: Intra-rater and inter-rater agreement in three different causality assessment methods

Causality assessment methods and comparison	No. of agreements (%) (N=401)	Agreement (Kappa) 95% C.I.	Level of agreement
WHO causality assessment scale	293 (73.07)	0.71 (0.65,0.77)	Substantial
Naranjo's algorithm	314 (78.30)	0.73 (0.66,0.77)	Substantial
European ABO	295 (73.57)	0.72 (0.68,0.76)	Substantial
Overall	270 (67.32)	0.69 (0.64,0.75)	Substantial

C.I.=Confidence interval

Research Letter

Table 2: Strengths and limitations of the three causality assessment methods

Causality assessment method	Strengths[2,3,11,12]	Limitations[2,3,11,12]
WHO-UMC method	Useful for the assessment of individual case reports Simple and easy to perform Universally accepted Adapted by Pharmacovigilance Programme of India for causality assessment	Causality depends on individual expertise and judgment Wide inter- and intra-observer disagreement Week reproducibility Individual categories are not sharply delineated
Naranjo's algorithm	Simple and brief. Most extensively used Method of scoring adds discerning value and reduces intra- and inter-observer variation Good reproducibility	In spontaneous reporting system, it lacks sensitivity because of many answers categorized as "unknown" in the absence of data. This results in spuriously low score for a clinically certain adverse drug reactions Dependability and validity not confirmed in children Not appropriate to be used in critically ill patients, organ toxicities and drug overdose
European ABO method	Easy to perform It combines documentation quality with causality Used in European Union to harmonize the reporting	Chances of overlap in causality assessment due to small number of categories leading to wrong interpretation of an adverse drug reaction Depends on individual expertise and judgment Intra- and inter-observer variability can be large Weak reproducibility

PvPI=Pharmacovigilance programme of India, UMC=Uppsala monitoring centre

rating as "probable" which is in consonance with Sriram *et al*.[10]

None of the causality assessment methods is either complete or reproducible. Very few methods take into consideration prior probability of the occurrence of an event with actual causal association. Also, categorization of causality into "possibly related" or "probably related" does not practically offer any additional advantage and leads to poor inter-rater agreement. The causality methods used in the present study have their own merits and demerits [Table 2]. We did find a good agreement between the three raters which could be the result of an actual causal association between the drug and the adverse event or due to individual expertise or commission of same error by all the raters simultaneously. However, the presence of confounding variables like underlying disease, concomitant medications, and absence of de-challenge/re-challenge information could have prevented the achievement of complete agreement.

Although the WHO-UMC scale is widely accepted and has been recommended by the National Pharmacovigilance Programme of India, we obtained the highest inter-rater agreement with Naranjo's algorithm. Additionally, it had a very good agreement with the other two methods as well. Hence, Naranjo's algorithm can be considered as a preferred method in individual case causality ascertainment because it is more objective and less dependent on personal expertise.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

Saket J. Thaker, Rahul S. Sinha, Nithya J. Gogtay, Urmila M. Thatte

Department of Clinical Pharmacology, Seth GS Medical College and KEM Hospital, Mumbai, Maharashtra, India

Address for correspondence:
Nithya Gogtay, 1st Floor, New MS Building,
Seth GS Medical College and KEM Hospital, Parel,
Mumbai - 400 012, Maharashtra, India.
E-mail: njgogtay@hotmail.com

REFERENCES

1. World Health Organization (WHO), Uppsala Monitoring Centre. The use of the WHO- UMC system for the standardized case causality assessment. WHO [online]. Available from: http://www.who-umc.org/graphics/4409.pdf. [Last accessed on 2014 Nov 14].
2. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: A systematic review. Drug Saf 2008;31:21-37.
3. 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.
4. Gallagher RM, Kirkham JJ, Mason JR, Bird KA, Williamson PR, Nunn AJ, *et al*. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. PLoS One 2011;6:e28096.
5. Fleiss Joseph L. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378-82.
6. Kappa index calculator. Available from: http://www.stattools.net/index.php. [Last accessed on 2013 Dec 26].
7. Davies EC, Rowe PH, James S, Nickless G, Ganguli A, Danjuma M, *et al*. An investigation of disagreement in causality assessment of adverse drug reactions. Pharm Med 2011;25:17-24.
8. Arimone Y, Miremont-Salamé G, Haramburu F, Molimard M, Moore N, Fourrier-Réglat A, *et al*. Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions. Br J Clin Pharmacol 2007;64:482-8.
9. Macedo AF, Marques FB, Ribeiro CF, Teixeira F. Causality assessment of adverse drug reactions: Comparison of the results obtained from

Research Letter

published decisional algorithms and from the evaluations of an expert panel. Pharmacoepidemiol Drug Saf 2005;14:885-90.

10. Sriram S, Ghasemi A, Ramasamy R, Devi M, Balasubramanian R, Ravi TK, *et al*. Prevalence of adverse drug reactions at a private tertiary care hospital in south India. J Res Med Sci 2011;16:16-25.

11. Seger D, Barker K, McNaughton C. Misuse of the naranjo adverse drug reaction probability scale in toxicology. Clin Toxicol (Phila) 2013;51:461-6.

12. Khan LM, Al-Harthi SE, Osman AM, AbdulSattar MA, Ali AS. Dilemmas of the causality assessment tools in the diagnosis of adverse drug reactions. Saudi Pharm J 2015. [Article in press].