International Journal of Research in Medical Sciences
Sharma S et al. Int J Res Med Sci. 2017 Oct;5(10):4389-4394
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20174564

Original Research Article

Inter-rater and intra-rater agreement in causality assessment of adverse drug reactions: a comparative study of WHO-UMC versus Naranjo scale

Sushil Sharma[1]*, Ajay Kumar Gupta[1], G. Jyothi Reddy[2]

[1]Department of Pharmacology, Armed Forces Medical College, (AFMC) Pune-40, Maharashtra, India
[2]Technical Associate (PvPI) ADR Monitoring Centre, AFMC Pune-40, Maharashtra, India

Received: 20 July 2017
Accepted: 18 August 2017

***Correspondence:**
Dr. Sushil Sharma,
E-mail: sushkehar@rediffmail.com

ABSTRACT

Background: The causality assessment system proposed by the WHO collaborating centre for international drug monitoring, the Uppsala monitoring centre (WHO-UMC) and the Naranjo probability scale are the generally accepted and most widely used methods for causality assessment. Both these scales are structured, transparent, consistent, and easy to apply with distinct advantages. The PvPI recommends use of WHO-UMC scale while many clinicians prefer Naranjo scale for its simplicity. As both these scales are used very widely in practise, it is important to study the level of agreement among them, to ensure proper interpretation of the causality assessment of the drugs. Objective was to study the inter-rater and intra-rater agreement in causality assessment of adverse drug reactions between WHO-UMC and Naranjo scale.
Methods: A total of 200 cases were analyzed by three raters. Each ADR reporting form was analyzed by the raters independently using both the Naranjo scale and the WHO-UMC scale. The inter-rater and Intra-rater agreement of all the three raters was analyzed using kappa statistics.
Results: The most common category of causality assessment was 'possible' which was around 73%, while the probable, definite and unlikely accounted for 23%, 3% and 1% respectively. The inter-rater agreement for the various categories of causality assessment when using Naranjo scale ranged from "very good to excellent" (Kappa value 0.95) while the same ranged from "good to very good" when the WHO-UMC scale was used (Kappa value 0.89) while the Intra-rater agreement for the three raters ranged from good to very good. The mean time (in minutes) taken for assessing the suspected adverse drug an ADR was 8. 26±2.53 for WHO-UMC scale while it was 14.18±3.44 when Naranjo scale was used.
Conclusions: Both the WHO-UMC scale and Naranjo scale are reliable and valid tools for causality assessment but the Inter-rater agreement was slightly better with Naranjo as compared to WHO-UMC scale.

Keywords: Causality assessment, Naranjo scale, Rater agreement, WHO-UMC scale

INTRODUCTION

The most common result of the meeting between the patient and the doctor is the prescription of drugs to a patient. Even after all aspects of "rational prescribing" are followed, the use of drugs by the patient leads to untoward and undesirable effects in many instances. These "adverse drug reactions" (ADRs) are recognised as a major cause of morbidity, hospital admission, and even death. The incidence of ADR's in hospitalized patients in

India has been reported as 6.34% while that of in Europe as 10.1% in the literature.[1,2] The monitoring of use of drugs and timely reporting of ADR's plays a very crucial role in the safe and effective use of drugs.

The pharmacovigilance programme of India (PvPI) initiated in 2010 has led to the establishment of ADR monitoring centres (AMC) in all medical colleges who report the ADR data collected to the national coordination centre (NCC) at Indian pharmacopoeia commission Ghaziabad. This information collected at the NCC becomes part of the national database which is then used to gain knowledge of the comprehensive safety profile of the drug so that appropriate actions can be taken in time to ensure minimal potential harm with the use of the drug.[3] As part of the analysis of the ADR data the staff also perform the 'causality assessment' to assess the causal relationship between a drug and an adverse event before reporting to NCC. "Causality assessment" assesses the relationship between a drug treatment and the occurrence of an adverse event. It is the cornerstone of Pharmacovigilance and contributes to better evaluation of the risk-benefit profiles of medicines and is an essential part of evaluating ADR reports in early warning systems and for regulatory purposes. Causality assessment can help in signal detection and aid in risk-benefit decisions regarding medicines.[3]

Many causality assessment methods have been proposed to assess the relationship between a drug and an adverse event in a given patient and the use of these standardized assessments for the relationship-likelihood of case reports of suspected ADRs is expected to provide a reliable reproducible measurement of causality.[4] The causality assessment system proposed by the world health organization collaborating centre for international drug monitoring, the Uppsala monitoring centre (WHO-UMC) and the Naranjo probability scale are the generally accepted and most widely used methods for causality assessment.[5,6] Both these scales are structured, transparent, consistent, and easy to apply assessment methods with distinct advantages. The PvPI recommends use of WHO-UMC scale while many clinicians prefer Naranjo scale for its simplicity. Due to the use of both these scales it is important to study the level of agreement among them, to ensure proper interpretation of the causality assessment of the drugs.

Many studies have been conducted on the issue and have generated varied data. Some studies reported Naranjo scale as better scale for causality assessment, while other studies mentioned WHO-UMC as better and simpler than Naranjo scale.[5,7,8]. But no study could conclude that a particular scale is ideal for causality assessment, and there were many discrepancies while comparing these scales.[9] Hence, the present study was planned.

METHODS

The study was conducted based on the ADR reports that were generated at the AMC of a medical college based on the suspected reactions from the affiliated teaching hospitals of the medical college. The reports were pertaining to the period from May 2014 to Dec 2015. The central drug standard control organization (CDSCO) ADR reporting forms were used for the collection of ADRs. The diagnosis of ADRs was primarily based on detailed histories and the correlation between drug intake and the onset of the ADR.

Table 1: Naranjo scale.[6,9,10]

Question	Yes	No	Don't know
Are there previous conclusion reports on this reaction?	+1	0	0
Did the adverse event appear after the suspect drug was administered?	+2	-1	0
Did the AR improve when the drug was discontinued or a specific antagonist was administered?	+1	0	0
Did the AR reappear when drug was re-administered?	+2	-1	0
Are there alternate causes (other than the drug) that could solely have caused the reaction?	-1	+2	0
Did the reaction reappear when a placebo was given?	-1	+1	0
Was the drug detected in the blood (or other fluids) in a concentration known to be toxic?	+1	0	0
Was the reaction more severe when the dose was increased or less severe when the dose was decreased?	+1	0	0
Did the patient have a similar reaction to the same or similar drugs in any previous exposure?	+1	0	0
Was the adverse event confirmed by objective evidence?	+1	0	0

Scoring for Naranjo algorithm: >9 = definite ADR; 5-8 = probable ADR; 1-4 = possible ADR; 0 = doubtful ADR.

A total of 200 cases collected during the study period were analyzed by three raters. The information for these 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 de-challenge. Each ADR reporting form was analyzed by the raters independently using both the Naranjo scale and the WHO-UMC scale. The inter-rater and intra-rater agreement of all the three raters was analyzed to study the agreement between them.

Table 2: WHO-UMC causality categories.[8]

WHO-UMC causality categories	
Certain	Event or laboratory test abnormality, with plausible time relationship to drug intake cannot be explained by disease or other drugs, Response to withdrawal plausible (pharmacologically, pathologically), Event definitive pharmacologically or phenomenologically (i.e., an objective and specific medical disorder or a recognized pharmacologic phenomenon), Rechallenge satisfactory, if necessary
Probable/ likely	Event or laboratory test abnormality, with reasonable time relationship to drug intake Unlikely to be attributed to disease or other drugs, Response to withdrawal clinically reasonable, Rechallenge not required
Possible	Event or laboratory test abnormality, with reasonable time relationship to drug intake could also be explained by disease or other drugs
Unlikely	Event or laboratory test abnormality, with a time to drug intake that makes a relationship improbable (but not impossible), Disease or other drugs provide plausible explanation
Conditional/ unclassified	Event or laboratory test abnormality, more data for proper assessment needed, or Additional data under examination
Unassessable/ unclassifiable	Report suggesting an adverse reaction, cannot be judged because information is insufficient or contradictory, Data cannot be supplemented or verified

Statistical analysis

The data management and analyses were performed using WIN PEPI version 11.65 software for inter-rater reliability while Graphpad Prism version 7.03 to assess the intra-rater agreement. We calculated agreement between three raters for each of Naranjo and WHO-UMC scale as well as agreement for each rater while using both of the scales for each ADR by generating Kappa coefficients with 95% confidence intervals. Kappa statistics represent the proportion of agreement greater than that expected by chance and are interpreted as represented ranging from Nil/Poor agreement to excellent agreement and are represented in Table 3.

Table 3: Interpretation of Kappa.[11,12]

Kappa value	Degree of agreement
≤ 0	Nil
0.01-0.20	Poor
0.21-0.40	Slight
0.41-0.60	Fair
0.61-0.80	Good
0.81-0.92	Very good
0.93-1.-00	Excellent

RESULTS

Out of the 200 suspected ADR's reported during the study period, most of the ADR's were reported from male patients (132 cases) while female patients accounted for a comparatively lesser number of adverse drug reactions reported (68 cases). The comparison of causality assessment categories by the three raters using Naranjo scale and WHO-UMC scale is represented in Figure 1 and Figure 2 respectively.

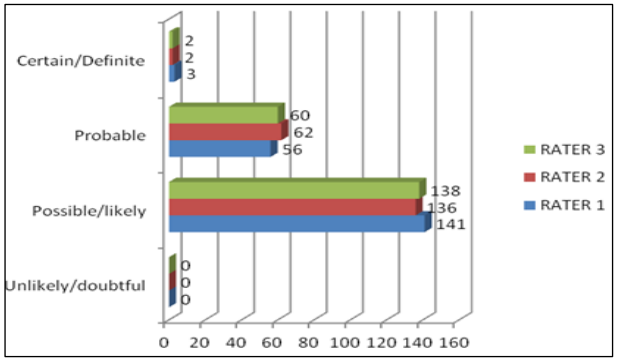

Figure 1: Causality assessment by the three raters: Naranjo scale.

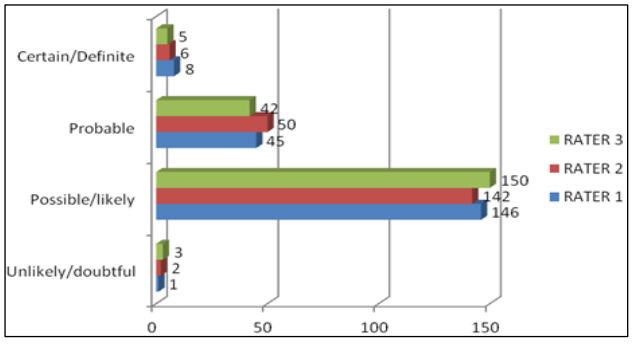

Figure 2: Causality assessment by the three raters: WHO-UMC scale.

The most common category of causality assessment was 'possible' which was around 73%, while the probable, definite and unlikely accounted for 23%, 3% and 1% respectively. As seen from the bar charts, the total number of causality attributes given by each rater are comparable for each category both when using the Naranjo scale and also when the WHO-UMC scale was used.

Further, the inter-rater agreement and also the intra-rater agreement for each of the adverse drug reaction using the two causality assessment scales was compared using the Kappa statistics. The inter-rater agreement for Naranjo scale is listed in Table 4, while inter-rater agreement for WHO-UMC scale is listed in Table 5.

Table 4: Inter-rater agreement between the three raters: Naranjo scale.

Category	Kappa	SE	Strength of agreement
Definite	0.86	0.041	Very good
Probable	0.95	0.041	Excellent
Possible	0.96	0.041	Excellent
Doubtful	1.0	0.0	Excellent

Table 5: Inter-rater agreement between the three raters: WHO-UMC scale.

Category	Kappa	SE	Strength of agreement
Definite	0.84	0.041	Very good
Probable	0.89	0.041	Very good
Possible	0.91	0.041	Very good
Doubtful	0.66	0.041	Good

The overall kappa test for inter-rater agreement using Naranjo scale was 0.94 with SE of 0.039, while using WHO-UMC scale was 0.83 with SE of 0.035, which indicates 'excellent' and 'very good' agreement. The intra-rater agreement using both the causality assessment scales was also studied and is listed in Table 6.

Table 6: Intra-rater agreements using both the scales: Naranjo and WHO-UMC scale.

	No of observed agreements	No of agreements expected by chance	Kappa	SE of Kappa	95% CI	Strength of agreement
Rater 1	188 (94%)	115.7 (57.83%)	0.858	0.038	0.783-0.932	Very good
Rater 2	186 (93%)	112.1 (56.06%)	0.841	0.039	0.764-0.918	Very good
Rater 3	179 (89.5%)	116.2 (58.07%)	0.750	0.049	0.653-0.846	Good

The intra-rater agreement for rater 1 was 'very good' with Kappa value of 0.858, SE of 0.038 and 95% confidence interval (CI) between 0.783-0.932.

The Kappa value for rater 2 was 0.841 with SE 0.039 and CI between 0.764-0.918, the agreement also being 'very good'. The intra-rater agreement for rater 3 was 'good', with Kappa value of 0.750, SE 0.049 and 95% CI 0.653-0.846. The mean time taken for assessing the suspected adverse drug an ADR was 8. 26±2.53 for WHO-UMC scale while it was 14.18±3.44 when Naranjo scale was used.

DISCUSSION

The decision to link an observed adverse reaction with a drug has to be taken with a great degree of responsibility. This decision if based solely on the clinical acumen and experience of the expert can lead to erroneous results and can compromise its validity and reliability.

Hence, over the years various scales have been developed to facilitate and guide the professionals in this decision-making process, the commonest among them being Naranjo scale and WHO-UMC scale. Various studies have been conducted to find out the suitability of one

method over the other. Son MK et al in 2008, evaluated 100 ADR cases using Naranjo scale and WHO-UMC scale and observed that the Naranjo probability scale was helpful for assessing unexpected ADRs and useful for evaluators with little experience.[7]

However, some of the items were not utilized and there were discrepancies when compared with the WHO-UMC causality criteria. The study by Belhekar M et al in 2014 assessed 913 ADRs collected between January 2010 and December 2012 using the WHO-UMC criteria and Naranjo algorithm and showed that there was a poor agreement between the WHO-UMC criteria and Naranjo algorithm, however it added that the WHO-UMC was less time-consuming.[5] Mittal N et al performed the casualty assessment using WHO and Naranjo scales on 200 ADR proformas.[9] The results of Kappa analysis demonstrated a moderate to good agreement between the two scales.

In the present study, a total of 200 cases collected during the study period were analysed by three trained raters who were sufficiently experienced in ADR reporting and causality assessment. All the three raters independently assessed the ADRs both by WHO-UMC and Naranjo scale. To ensure that confounding parameters relating to quality of information in the ADR reporting form are addressed, all the three raters assessed the same ADR's and had similar information available to them on which to base their causality assessment.

In the present study, the inter-rater agreement while using Naranjo scale was better as compared to WHO-UMC scale. The Inter-rater agreement for the various categories of causality assessment when using Naranjo scale ranged from "very good to excellent" while the same ranged from "good to very good" when the WHO-UMC scale was used. The better agreement with Naranjo scale between the various raters can be attributed to the fact that, in Naranjo scale, the assessment is being done by using a specific set of questions, with answers 'yes', 'no' or 'don't know' with specific scores assigned to each answer. This is different in WHO-UMC scale, in which all points given for each category should be reasonably complied with to assess the particular ADR. This may lead to increased subjectivity with higher chances of inter-rater variations with WHO-UMC as compared to Naranjo scale.

The time taken for causality assessment by WHO-UMC scale was lesser as compared to Naranjo scale which is in agreement with earlier studies.[14] This is probably because as mentioned above the Naranjo scale has specific questions that need to be answered before score is allotted to the questions and the rater has to be very clear and objective for each of the questions which is time consuming. This is in contrast to the WHO-UMC scale where some amount of subjectivity can creep and consequently lesser time may be spent by the rater before ascribing causality. However, the intra-rater agreement between the two scales ranged from "good to very good" which shows showing that both the scales are equally reliable as tools for causality assessment.

CONCLUSION

From the present study, it has been concluded that both WHO-UMC scale and Naranjo scale were reliable and valid tools for causality assessment but the Inter-rater agreement was slightly better with Naranjo as compared to WHO-UMC scale.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Patel TK, Patel PB. Incidence of adverse drug reactions in Indian hospitals: a systematic review of prospective studies. Curr Drug Saf United Arab Emirates. 2016;11(2):128-36.
2. Bouvy JC, De Bruin ML, Koopmanschap MA. Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. Drug Saf New Zealand. 2015;38(5):437-53.
3. Pharmacovigilance programme of India (Internet). Performance Report 2015-16. 2016. Available at http://www.ipc.gov.in/PvPI/pub/Pharmacovigilance%20Programme%20of%20India%20Annual%20performance%20Report%202015-2016.pdf.
4. Agbabiaka TB, Savovic J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. Drug Saf. 2008;31(1):21-37.
5. Belhekar MN, Taur SR, Munshi RP. A study of agreement between the Naranjo algorithm and WHO-UMC criteria for causality assessment of adverse drug reactions. Indian J Pharmacol. 2014;46(1):117-20.
6. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther. 1981;30(2):239-45.
7. Son MK, Lee YW, Jung HY, Yi SW, Lee KH, Kim SU, et al. Comparison of the Naranjo and WHO-Uppsala monitoring centre criteria for causality assessment of adverse drug reactions. Korean J Med. 2008;74:181-7.
8. Rehan HS, Chopra D, Kakkar AK. Causality assessment of spontaneously reported adverse drug events: comparison of WHO-UMC criteria and Naranjo probability scale. Int J Risk Saf Med. 2007;19:223-7.
9. Mittal N, Gupta MC. Comparison of agreement and rational uses of the WHO and Naranjo adverse event causality assessment tools. J Pharmacol Pharmacother. 2015;6:91-3.
10. García-Cortes M, Lucena MI, Pachkoria K, Borraz Y, Hidalgo R, Andrade RJ. Evaluation of Naranjo

adverse drug reactions probability scale in causality assessment of drug-induced liver injury. Aliment Pharmacol Ther. 2008;27(9):780-9.
11. Byrt T. How good is that agreement? Epidemiol. 1996;7:561.
12. McHugh M. Interrater reliability: the kappa statistic. Biochem Medica. 2012;22(3):276-82.
13. Viera AJ, Garrett JM. Understanding interobserver agreement: The Kappa statistic. Fam Med. 2005;37(5):360-3.
14. Sharma M, Gupta SK, Gupta VB, Chatterjee A. A comparative study of causality assessment scales used in the analysis of spontaneously reported events: WHO-UMC criteria versus Naranjo probability scale. J Pharmacovig Drug Safety. 2009;6(1):5-9.

Cite this article as: Sharma S, Gupta AK, Reddy GJ. Inter-rater and intra-rater agreement in causality assessment of adverse drug reactions: a comparative study of WHO-UMC versus Naranjo scale. Int J Res Med Sci 2017;5:4389-94.