Original Article

Agreement between WHO-UMC causality scale and the Naranjo algorithm for causality assessment of adverse drug reactions

Ajay K. Shukla[1], Ratinder Jhaj[1], Saurav Misra[1], Shah N. Ahmed[2], Malaya Nanda[1], Deepa Chaudhary[3]

[1]Department of Pharmacology, AIIMS Bhopal, Madhya Pradesh, [2]Department of Pharmacology, College of Medicine and JNM Hospital, Kalyani, West Bengal, [3]Department of ADR Monitoring Centre (AMC) AIIMS Bhopal, AIIMS Bhopal, Madhya Pradesh, India

Abstract

Background: The Pharmacovigilance Program of India recommends the use of the World Health Organization-Uppsala Monitoring Centre (WHO-UMC) scale, while many clinicians prefer the Naranjo algorithm for its simplicity. In the present study, we assessed agreement between the two widely used causality assessment scales, that is, the WHO-UMC criteria and the Naranjo algorithm. **Materials and Methods:** In this study, 842 individual case safety reports were randomly selected from 1000 spontaneously reported forms submitted to the ADR Monitoring Center at a tertiary healthcare Institute in Central India between 2016 and 2018. Two well-trained independent groups performed the causality assessment. One group performed a causality assessment of the 842 ADRs using the WHO-UMC criteria and the other group performed the same using the Naranjo algorithm. The agreement between two ADR causality scales was assessed using the weighted kappa (κ) test. **Results:** Cohen's kappa coefficient (κ) statistical test was applied between the two scales (WHO-UMC scale and Naranjo algorithm) to find out the agreement between these two scales. "No" agreement was found between the two scales {Kappa statistic with 95% confidence interval = 0.048 (*P* < 0.001)}. **Conclusion:** There was no agreement found between the WHO-UMC criteria and the Naranjo algorithm in our study.

Keywords: Agreement between scales, causality assessment, Naranjo algorithm, WHO-UMC scale

Introduction

An adverse drug reaction (ADR), according to World Health Organization (WHO), is a "response to a drug that is noxious and unintended and that occurs at doses used in humans for prophylaxis, diagnosis, or therapy of disease or for the modification of physiologic function".[1] The incidence of ADR reported by various studies across the world is 6–20%, whereas in India, it is up to 3%. About 10–20% ADRs reported are from hospitalized patients, which leads to prolongation of a stay.[1,2] Assessment of an ADR is an integral component of the definition of pharmacovigilance. It is the connecting link between the detection, understanding, and prevention of ADRs. It assesses the strength of association of the ADR with the suspected drug to evaluate whether it can be further processed for signal generation. A causality assessment method, which is valid, reliable, and universally acceptable, has been an elusive target in pharmacovigilance.[3]

Many factors can contribute to the occurrence of an ADR, like patient-related factors, drug-related factors, and disease-related

Access this article online

Quick Response Code:

Website:
www.jfmpc.com

DOI:
10.4103/jfmpc.jfmpc_831_21

How to cite this article: Shukla AK, Jhaj R, Misra S, Ahmed SN, Nanda M, Chaudhary D. Agreement between WHO-UMC causality scale and the Naranjo algorithm for causality assessment of adverse drug reactions. J Family Med Prim Care 2021;10:3303-8.

factors.[4] Causality assessment of ADRs is performed by clinicians, academics, the pharmaceutical industry, and regulators and in different settings, including clinical trials.[5,6] Thus, the assessment helps evaluate the risk-benefit profiles of medicines and is an essential part of assessing ADR reports as early warning systems and for regulatory purposes.[6,7]

Causality assessment methods can be broadly classified into three categories: global introspection, algorithms, and probabilistic methods.[8] Global introspection methods help in identifying the relationship likelihood in actual clinical practice. One standard method under this category is the World Health Organization-Uppsala Monitoring Centre (WHO-UMC) scale [Table 1].[7] The second category of causality assessment methods consists of algorithms that are simple and mostly questionnaire-based methods in which scores are assigned to answers to its questions.[8] Algorithms are inflexible, designed to reduce intrarater and interrater variability to make them more reliable and valid. The most commonly available criteria used worldwide are the Naranjo algorithm [Table 2a and 2b].[9,10] The third category of causality assessment methods is the probabilistic methods. These analyze a probability for causality to be calculated from available knowledge (previous estimation) and the specific findings in ADR reports, in combination with the background information (posterior estimate).[8]

Most Individual Safety Case Reports (ICSRs) are reported as suspected ADRs. As most of the ADRs are not specific for a particular drug and with absent diagnostic tests and a rechallenge ethically unjustified, causality assessment of an ADR becomes crucial. None of the current causality assessment methods produces a reliable and precise quantitative estimation of relationship likelihood.[7]

The causality assessment of an ADR has also numerous limitations. These limitations include their inability to provide an accurate measure of relationship likelihood, to distinguish valid from invalid cases, to prove the association between drug and event, to quantify the contribution of a drug to the adverse event development, and to change uncertainty into certainty.[7]

WHO-UMC scale is the causality assessment used by WHO Programme for International Drug Monitoring (PIDM). WHO-UMC scale [Table 1], based on clinical pharmacology knowledge, is widely used under the Pharmacovigilance Programme of India (PvPI).[7,8] WHO-UMC scale was developed as a practical tool for the causality assessment of ADRs as a part of the PIDM and as a practical tool for the assessment of case reports. It is a composite assessment method that takes into account the clinical–pharmacological aspects and the quality of the documentation of the ICSR.[7]

Naranjo algorithm is another simple widely used causality assessment method. Naranjo algorithm was developed to standardize the causality assessment of ADRs. This algorithm

can not only be applied in routine clinical practice but also in controlled trials of new medications. Nevertheless, it is simple to apply and widely used. Many publications on drug-induced liver injury mention the results of applying the ADR probability scale.[11]

There are many causality assessment tools (CATs), most commonly used are WHO-UMC criteria and the Naranjo algorithm. Currently, none of the CATs have been universally accepted as the gold standard. The PvPI recommends the use of the WHO-UMC scale, while many primary care physicians prefer the Naranjo algorithm for its simplicity. Poor reproducibility and varying levels of agreement have been observed among different CATs in assessing ADRs. Hence, the present was conducted to examine the agreement among two different CATs in assessing ADRs.

Materials and Methods

Study site

Department of Pharmacology, All India Institute of Medical Sciences Bhopal, which is a Regional Training Center and ADR monitoring center under the PvPI.

Data Collection: Complete ICSRs concerning all the required information were randomly selected from 1000 spontaneously reported forms submitted to the study site. To avoid observer bias, a causality assessment was performed by two independent groups. The institutional pharmacovigilance committee performed the causality assessment using the WHO-UMC scale [Table 1], and two clinical pharmacologists performed the causality assessment using the Naranjo algorithm [Table 2a and 2b]. Causality assessment of ADRs reported in ICSRs obtained with WHO-UMC criteria was categorized into certain, probable, possible, unlikely, unclassified, and unclassifiable. Similarly, in the Naranjo algorithm, ADRs were categorized into definite, probable, possible, and doubtful. The raters neither had direct medical or personal contact with the patients involved nor had any access to the patients' files.

Sample size calculation

To test the null hypothesis of H0: k0 = 0.1 against the alternative hypothesis HA: k > 0.1 with a significance level (α) of 0.05 and power (1-β) of 0.90 for k1 = 0.21 (from previous studies), using the formula given by Cantor,[12] the calculated minimum required sample size for the study was 830 ADR cases.

Statistics

Data were expressed as proportions or percentages of total observations. The agreement between two ADR causality scales was assessed using the weighted kappa (κ) test. Cohen's kappa coefficient (κ) statistic was used to measure interrater reliability (and also intrarater reliability) for qualitative (categorical) items. It is one of the more robust measures than simple percent agreement calculation, as κ takes into account the possibility of

Table 1: World Health Organization-Uppsala Monitoring Center (WHO-UMC) causality categories	
Causality term	**Assessment criteria (all points should be reasonably complied)**
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
	Information on drug withdrawal may be lacking or unclear
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

Table 2a: Naranjo Algorithm - ADR probability scale			
Question	**Yes**	**No**	**Don't know**
1. Are there previous conclusion reports on this reaction?	1	0	0
2. Did the adverse event appear after the suspect drug was administered?	2	-1	0
3. Did the AR improve when the drug was discontinued or a specific antagonist was administered?	1	0	0
4. Did the AR reappear when drug was readministered?	2	-1	0
5. Are there alternate causes [other than the drug] that could solely have caused the reaction?	-1	2	0
6. Did the reaction reappear when a placebo was given?	-1	1	0
7. Was the drug detected in the blood [or other fluids] in a concentration known to be toxic?	1	0	0
8. Was the reaction more severe when the dose was increased or less severe when the dose was decreased?	1	0	0
9. Did the patient have a similar reaction to the same or similar drugs in any previous exposure?	1	0	0
10. Was the adverse event confirmed by objective evidence?	1	0	0

Table 2b: Naranjo Algorithm - Interpretation of scores	
Score	**Interpretation of Scores**
Total Score ≥9	Definite. The reaction (1) followed a reasonable temporal sequence after a drug or in which a toxic drug level had been established in body fluids or tissues, (2) followed a recognized response to the suspected drug, and (3) was confirmed by improvement on withdrawing the drug and reappeared on re-exposure
Total Score 5-8	Probable. The reaction (1) followed a reasonable temporal sequence after a drug, (2) followed a recognized response to the suspected drug, (3) was confirmed by withdrawal but not by exposure to the drug, and (4) could not be reasonably explained by the known characteristics of the patient's clinical state
Total Score 1-4	Possible. The reaction (1) followed a temporal sequence after a drug, (2) possibly followed a recognized pattern to the suspected drug, and (3) could be explained by characteristics of the patient's disease
Total Score ≤0	Doubtful. The reaction was likely related to factors other than a drug

the agreement occurring by chance. 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. The κ value ranges from − 1 (perfect disagreement) to + 1 (perfect agreement). Statistical analysis was performed using Graph Pad Quick Calcs software available online at http://graphpad.com/quickcalcs.

Results

Out of 842 ADRs assessed in the study, 517 (61.4%) were seen in male patients and 325 (38.59%) in female patients. The most frequently assigned causality category in the Naranjo algorithm was "probable" (75.05%), while in the WHO-UMC scale, it was "certain" (63.33%) [Figures 1 and 2]. In the WHO-UMC scale, the second most common category was probable (in 20.4%)
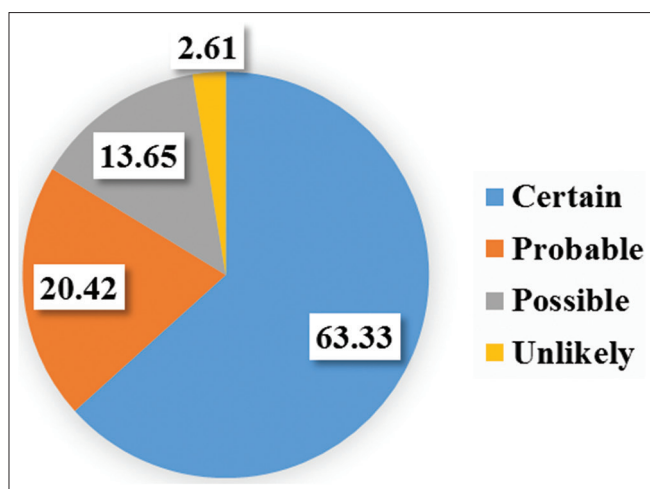
Figure 1: Distribution of sampled adverse drug reactions (ADRs) based on causality reported as per the WHO-UMC scale (%)

Figure 2: Distribution of sampled adverse drug reactions (ADRs) based on causality reported as per the Naranjo algorithm

followed by possible (in 13.6%) and unlikely (in 2.61%). In the Naranjo algorithm, the second most common category was possible (in 24.82%) followed by definite (in 0.11%) while there was no patient with doubtful category, as shown in Figure 2. There was "no" agreement found between the two scales {Kappa statistic with 95% confidence interval = 0.048 ($P < 0.001$)}.

Discussion

In this study, 842 ADRs were analyzed using both the Naranjo algorithm and the WHO-UMC scale by two different teams of assessors. The agreement between the two ADR causality scales was assessed using the weighted kappa (κ) test. On applying Cohen's kappa coefficient (κ) statistical test between the two scales (WHO-UMC scale and Naranjo algorithm) to find out the agreement between these two scales. There was "no" agreement found between the two scales (Kappa statistic with 95% confidence interval = 0.048 ($P < 0.001$). Thus, there was no agreement between the WHO-UMC scale and the Naranjo algorithm for 842 ADRs assessed for the causality.

The findings of the study are in congruence with the study by Belhekar M *et al.*,[13] in which 913 ADRs were assessed using the WHO-UMC criteria and Naranjo algorithm. They found poor agreement between these two scales (Kappa statistic with 95% confidence interval = 0.143), However, WHO-UMC was found to be less time consuming. Rana *et al.*[14] also found poor agreement (Cohen's κ =0.014) between the WHO-UMC scale and the Naranjo algorithm. In this study, only 36 ADRs were analyzed, all belonging to the pediatric age group.

There have been many studies in which the levels of agreement were found to be more than the levels of agreement found in this study. Behera *et al.*[15] did the causality assessment by the same team by using the WHO-UMC scale and the Naranjo algorithm. They found fair agreement (Cohen's κ =0.45) between the WHO-UMC system and the Naranjo algorithm. The difference in the levels of agreement found by Behera *et al.*[15] and this study can
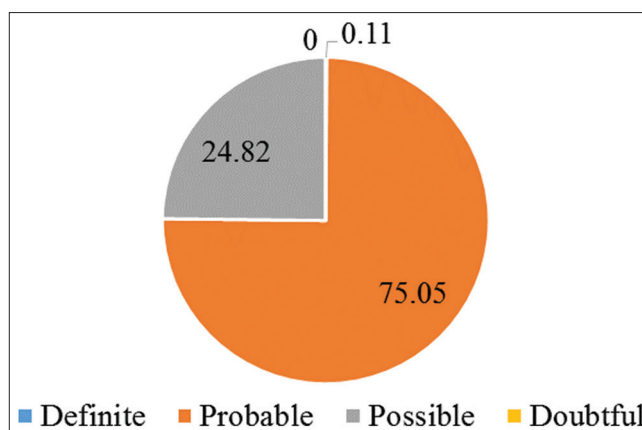
Table 3: Interpretation of Cohen's kappa		
Value of Cohen's kappa	Level of agreement	% of data that are reliable
0-0.20	None	0-4%
0.21-0.39	Minimal	5-15%
0.40-0.59	Weak	16-35%
0.60-0.79	Moderate	36-63%
0.80-0.90	Strong	64-81%
Above 0.90	Almost perfect	82-100%

be attributed to the smaller sample size (239 ADR reports) and observer bias as the same assessors did the causality assessment by both methods.[16]

Mittal *et al.*[17] found moderate agreement (Cohen's κ =0.701) between the WHO-UMC system and the Naranjo algorithm. Mittal *et al.*[17] did the causality assessment by two independent pharmacologists by applying both the methods and removed any discrepancy by mutual discussion. Acharya *et al.*[18] found moderate agreement (Cohen's κ =0.6) between the WHO-UMC system and the Naranjo algorithm. In this study, sample size was only 59 and both assessments were done by the same teams. The difference in the values found in the study done by Mittal *et al.*[17] and Acharya e*t al.*,[18] and this study can be attributed to the smaller sample size of the study (200) and observer bias. The study done by Rehan *et al.*[19] did the causality assessment of 1339 ADRs using the WHO-UMC scale and Naranjo algorithm by the same assessors and subsequently analyzed for agreement. In this study, they found moderate agreement (Cohen's κ =0.669).

Goyal *et al.*[20] assessed 1229 ADRs belonging to antihypertensive drugs using the WHO-UMC method, Naranjo Algorithm, and Versatile Causality Assessment Tool (VCAT method). WHO-UMC method and Naranjo algorithm had a good agreement (κ =0.669). In this study to avoid observer bias, ADRs were assessed by the same assessor at 3 months gap after assessing them by one scale. The level of agreement found by Goyal *et al.*[20] was more than that of our study. One of the possible reasons for the difference in the strength of agreement in the study done by Goyal *et al.* with

our study was that all the sampled suspected ADRs belonged to only one class of drug (antihypertensives).

Sharma *et al*.[21] analyzed 200 ADRs by three raters independently using both the Naranjo algorithm and the WHO-UMC scale. The interrater and intrarater agreement of all the three raters was analyzed to study the agreement between them. There was a "very good" agreement between both the Naranjo algorithm and the WHO-UMC scale (Kappa statistic with 95% confidence interval = 0.94). One of the possible reasons for the observed difference in the findings can be the different distribution of categories of sampled ADRs as per the WHO-UMC scale. In this study, the most common category of causality assessment was possible (73% ADRs) followed by probable, definite, and unlikely which accounted for 23, 3, and 1% of ADRs, respectively. Observed differences between our study and this study could also be due to subjective assessment of methods of ADR assessment by three different raters in the study, a small number of ADRs (200) form assessed, and interrater differences and completeness of the information. Apart from this, both the criteria have limitations and issues like mandatory rechallenge for certainty in WHO-UMC and subjectivity of interpretation in questions of Naranjo algorithm.

Son MK *et al*.[22] assessed 100 ADRs using the WHO-UMC criteria and Naranjo algorithm. The Spearman rank coefficient was 0.519 ($P < 0.001$), and the agreement was 55% between the Naranjo probability scale and the WHO-UMC causality categories. This difference in the study results could be due to different statistical tools used and smaller sample size.

In most of these studies, causality assessments from both scales were done by the same assessors. While this helps to reduce interrater variability in assessment, it can lead to observer bias. Observer bias is a type of systematic discrepancy from the true value during the observation and recording of data and plays an important role in studies; the observer has to use judgment to decide what is to be recorded.[16] Observer bias in the studies can result in overestimation of the degree of agreement between the two causality assessment scales. In this study, assessment by WHO-UMC scale and Naranjo algorithm was done by two independent teams to avoid observer bias.

There are differences between the WHO-UMC scale and the Naranjo algorithm in the range of questions asked which may influence the outcome. Each causality assessment method has its pros and cons. The selection of a particular causality assessment method and its interpretation is influenced by the availability of manpower, facilities, and knowledge of the assessor.

The poor agreement between these two scales found in this study can also be attributed to the different nature of the WHO-UMC scale and the Naranjo algorithm. WHO-UMC is one of the global introspection methods, which is nonprobabilistic and has unpredictability in evaluation. It has an inherent tendency to have subjective variations in causality assessment because of differences in the knowledge and expertise of clinicians. Naranjo

algorithms are designed to reduce interrater and intrarater dissimilarity. It has low sensitivity but reasonable specificity. Furthermore, algorithms improvise the logical feature of causality assessment, and they are frequently employed to spot ICSRs. It cannot consistently ascertain the causality due to the lack of regard to the "confounding variables" like underlying disease, concurrent use of other drugs, and lack of available ADRs knowledge.

Another important feature of causality assessment is the average time taken for the causality assessment. Time taken using the WHO-UMC method was comparatively lesser than the Naranjo algorithm as also presented in studies by Rehan e*t al*.[13] and Belhekar e*t al*.[13] As mentioned above, the Naranjo algorithm has specific questions that require the rater to be objective for each of the questions which can be time consuming. This finding is in contrast to the WHO-UMC scale, where the interpretation of subjectivity can creep, and consequently, lesser time may be spent by the rater before deciding causality.

Limitations – One potential factor that may have contributed to the low level of agreement is the interrater variability between the two teams of assessors. Subgroup analyses such as comparing ADRs obtained from children versus adults, acute versus chronic ADRs, and suspected versus unsuspected ADRs were not done.

However, this study is significant as sample size calculation was done beforehand for adequate power of the study. Observer bias was avoided by using independent teams for causality assessment by different scales.

Conclusion

There was no agreement found between the WHO-UMC criteria and the Naranjo algorithm in this study. Standardization of CAT for the development of universally acceptable causality assessment method is the need of the hour. As causality assessment plays a vital role in the pharmacovigilance process, there is a need for developing a universally acceptable objective causality assessment scale. Future studies can be planned where interrater variability can be assessed using the same scales. Further studies are also required to develop the gold standard method for the causality assessment of ADRs.

Summary

- There was no agreement found between the WHO-UMC criteria and the Naranjo algorithm in this study.
- As causality assessment plays a vital role in the pharmacovigilance process, there is a need for developing a universally acceptable objective causality assessment scale.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

References

1. Mulchandani R, Kakkar AK. Reporting of adverse drug reactions in India: A review of the current scenario, obstacles and possible solutions. Int J Risk Saf Med 2019;30:33–44.

2. Singh P, Agrawal M, Hishikar R, Joshi U, Maheshwari B, Halwai A. Adverse drug reactions at adverse drug reaction monitoring center in Raipur: Analysis of spontaneous reports during 1 year. Indian J Pharmacol 2017;49:432-7.

3. World Health Organization. Safety of medicines. A guide to detecting and reporting adverse drug reactions – why health professionals need to take action. Geneva: World Health Organization; 2002.

4. Alomar MJ. Factors affecting the development of adverse drug reactions. Saudi Pharm J 2014;22:83-94.

5. Agbabiaka TB, Savovicx J, Ernst E. Methods for causality assessment of adverse drug reactions: A systematic review. Drug Saf 2008;31:21–37.

6. Turner WM. The Food and Drug Administration algorithm. Special workshop: Regulatory. Drug Inform J 1984;18:259–66.

7. World Health Organization (WHO)-Uppsala Monitoring Centre. The use of the WHO-UMC system for standardized case causality assessment. Available from: http://www.who-umc.org/Graphics/24734.pdf.

8. Theophile H, Arimone Y, Miremont SG, Moore N, Fourrier RA, Haramburu B. Comparison of three methods (consensual expert judgment, algorithmic and probabilistic approaches) of causality assessment of adverse drug reactions: An assessment using reports made to a French Pharmacovigilance center. Drug Saf 2010;33:1045-54.

9. 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;30:239–45.

10. Mashford ML. The Australian method of drug-event assessment. Special workshop – Regulatory. Drug Inf J 1984;18:271–3.

11. LiverTox: Clinical and Research Information on drug-induced liver injury. Adverse drug reaction probability scale (Naranjo) in drug induced liver injury. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; 2012. Available from: https://www.ncbi.nlm.nih.gov/books/NBK548069/. [Updated 2019 May 04].

12. Cantor AB. Sample-size calculations for Cohen's Kappa. Psychol Methods 1996;1:150–3.

13. 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;46:117-20.

14. Rana DA, Bhadiyadara SN, Shah HJ, Malhotra SD, Patel VJ. Consistency between causality assessments obtained with various scales and their agreement for adverse drug events reported in pediatric population. J Young Pharm 2015;7:89–95.

15. Behera SK, Das S, Xavier AS, Velupula S, Sandhiya S. Comparison of different methods for causality assessment of adverse drug reactions. Int J Clin Pharm 2018;40:903-10.

16. Mahtani K, Spencer EA, Brassey J, Heneghan C. Catalogue of bias: Observer bias. BMJ Evid Based Med 2018;23:23–4.

17. 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.

18. Acharya TA, Trivedi MD, Joshi KJ, Chhaiya SB, Mehta DS. A study of agreement between WHO-UMC causality assessment system and the Naranjo algorithm for causality assessment of adverse drug reactions observed in medical ICU of a tertiary care teaching hospital. Biomed Pharmacol J 2020;13.

19. 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.

20. Goyal R, Sodhi RK, Gupta A. Comparison of three methods of causality assessment of adverse drug reactions to antihypertensive drugs. Int Res J Pharm 2019;10:195-9.

21. 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.

22. 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.