

Causality assessment of adverse drug reactions detected in a Brazilian public hospital: interrater agreement using a global introspection method

Renata Menezes¹, Maria das Dores Silva¹, Antonio Ribeiro¹, Marcelo Filho¹, Glaucia Martinho¹, Luna Ferreira², and Maria Martins³

¹Hospital das Clinicas da Universidade Federal de Minas Gerais

²Fundação Hospitalar do Estado de Minas Gerais

³Universidade Federal de Minas Gerais Faculdade de Medicina

May 6, 2020

Abstract

Adequate categorization of adverse drug reaction (ADR) causality may contribute to its early recognition in the assessment of ADR. The method proposed by World Health Organization and Uppsala Monitoring Center (WHO-UMC system) is universally used for causality assessment. We sought to evaluate ADR causality in hospitalized patients using WHO-UMC system. ADR preventability was also evaluated in this study. We reviewed cases of ADR identified in a Brazilian university hospital. Three healthcare professionals adjudicated, independently, each ADR reported using WHO-UMC method. Overall, 90 ADR involving 64 patients were evaluated and adjudicated as possible in 47 (52.2%) cases. Slight concordance was found in the agreement analyses, with kappa range of 0.10-1.18, indicating poor reproducibility of WHO-UMC system. Our results suggest the involvement of interdisciplinary teams could improve the quality of the monitoring process of ADR. A great number of preventable ADR reinforce the need for improvements in pharmacovigilance strategies in hospital settings.

Introduction

The causality assessment of adverse drug reactions (ADR) is an essential approach in pharmacovigilance, as an attempt to investigate the connection between the use of a drug and a suspected ADR. WHO-UMC system is a universally used method for causality assessment.[1] Previous data emphasized its high consistency to be applied in hospitalized patients. However, there are some limitations raised by some authors regarding the fact that it is based on individual knowledge and judgment, and a poor reproducibility has been encountered among different judges.[1, 2] Naranjo algorithm is another method used to assess ADR causality which has been reported to be simple and brief. This is widely used due to its easy implementation and high interrater agreement, despite poor sensitivity.[1-4]

Causality assessment presupposes theoretical-methodological integration, but requires judgments which are essential for drug safety evaluation.[4] Adequate categorization of causality, especially in highly complex institutions, may contribute to the early recognition of ADR and their consequent prevention, in addition to the optimization of drug therapy.[3] We sought to assess causality of ADR in patients hospitalized in a university hospital, using the WHO-UMC method, to evaluate the performance of Naranjo algorithm[5] and ADR preventability using an adapted Schumock and Thornton method.[6, 7]

Methods

We reviewed cases of ADR previously identified in a study including 300 adult inpatients. These patients were recruited from January 26 to March 6, 2015 in surgical and medical wards of a general 500-bed university

hospital located in Belo Horizonte, Southeastern Brazil. The study was approved by the Research Ethics Committee of the Universidade Federal de Minas Gerais. All participants signed a patient consent statement. In the present study, ADR was defined as “a response to a medicinal product which is noxious and unintended and which occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease or for the restoration, correction or modification of physiological function”, according to WHO.[8, 9] The methods used to identify ADR were described elsewhere.[9]

Three healthcare professionals acted as judges to adjudicate, independently, each ADR reported in the study according causality categories of ADR. The professionals had experience in assistance and research. They had different academic background, including a nurse (Judge A), a pharmacist (Judge B) and a physician (Judge C). Standardized forms were evaluated by the judges and included the following information: patient’s data (age, sex, diagnosis, type of comorbidities, laboratory data, daily clinical data and drug therapy) and ADR data (onset/end of clinical manifestation, history of allergies, administered doses, drug interactions, suspected drugs to cause ADR and time of its use).

Expert adjudication process consisted in the evaluation of cases by each judge, separately, followed by the record of his/her judgment in an instrument developed specifically for this study. ADR were classified into six categories (certain, probable, possible, unlikely, conditional, unassessable), according to WHO-UMC criteria and based on individual expertise. Additionally, the judge was required to register the suspected drug(s) to have caused ADR.

ADR causality was also evaluated using Naranjo algorithm which is a method consisting in ten yes/no questions.[5] They were evaluated in this study consensually by two clinical pharmacists with pharmacovigilance expertise. Preventability of ADR was assessed by the judges, using Schumock and Thornton method, adapted from Shajahan *et al.* [6, 7]

Statistical analysis

Descriptive statistics were employed to determine numerical and categorical variables. Numerical variables were presented using measures of central tendency and dispersion (mean, median, standard deviation (SD) and interquartile range (IQR), as appropriate). Absolute and relative frequencies were presented for categorical variables. Overall interrater agreement, among three judges, was calculated using Fleiss’ kappa coefficient.[10] Cohen’s kappa coefficient with linear weighting[11] was used to measure pairwise judge concordance. For both kappa coefficient measures, all cases in which ADR were classified as unassessable by at least one judge were excluded. We considered 95% of confidence interval for all the analyses. The assessment of concordance was based on the quantitative scale proposed by Landis and Koch[12]: <0=poor, 0.00-0.20=slight, 0.21-0.40=fair, 0.41-0.60=moderated, 0.61-0.80=substantial, 0.81-0.99=almost perfect and 1.00=perfect.

Results

A total of 90 ADR involving 64 patients were included in the analyses. Patients presented a median age of 63 years (IQR: 51.0-70.8) and were predominantly male (51.6%). Descriptive data and agreement analyses of ADR were presented in **Table 1**. Twelve different types of clinical manifestations were found, resulting from 30 suspected drugs to cause ADR. Most patients presented only one ADR (n=45; 70.3%). Thirteen patients presented two ADR (20.3%) and five patients presented three ADR (7.8%) during the follow up. Maximal number of ADR per patient was four, identified in only one patient (1.6%). Mean ADR duration was three days, and 13 ADR (14.4%) lasted for more than five days.

A total of 270 assessments were performed using the WHO-UMC system, as a result of the evaluation of 90 ADR by the three judges involved in the adjudication process. According to global (majority) classification, 47 (52.2%) ADR were categorized as possible and 37 (41.1%) as probable. Conditional and unlikely were classified in three (3.3%) and two (2.2%) cases, respectively. Only one ADR was classified as definite. Similar categorization of the ADR, according to Naranjo algorithm was also presented in Table 1.

Slight agreement was found in the comparison of all judges in the pairwise analysis, and also for multiple

judges, indicating poor reproducibility of WHO-UMC system (**Table 1**) . Thirteen cases were classified as unassessable by at least one judge and they were excluded for kappa analysis.

More than a half of the studied cases of ADR were categorized by the judges with some level of preventability. ADR was considered definitely preventable in 29 (32.2%) cases and probably preventable in 23 (25.6%) cases. Thirteen eight (42.2%) cases of ADR were considered not preventable by the participating judges. Kappa Fleiss also revealed a slight concordance among the judges ($k=0.134$; CI 95%: 0.018-0.225).

Discussion

Our study evaluated the adjudication processes for the causality assessment of ADR and the agreement between WHO-UMC. Data suggested slight interrater agreement for WHO-UMC, according to Lands and Koch scale.[12] Previous studies have reported interrater agreement ranging from fair to substantial agreement in the ADR causality assessment involving inpatients.[3, 13] In hospitalized patients with high complexity of clinical condition, this classification is particularly difficult, since many factors are involved in the scenario, such as polypharmacy, multiple comorbidities, different healthcare teams with diverse clinical experience involved in patient care.

WHO-UMC system is based on standardized criteria, demanding minimal expertise in pharmacovigilance process, although clinical practice is an important requisite for judges. This dependence on the individual judgement and different judges' background could also explain the expected weak reproducibility and slight interrater agreement, as cited by Khan *et al.*, [4] and demonstrated by our findings. Nevertheless, from the perspective of patient safety, the evaluation performed by a multiprofessional team could enrich the assessment of ADR given the diverse experiences and subjectivity. In the field of pharmacovigilance, the search for consensus among professionals is highly desirable to improve the quality of the process instead of prioritizing their individual view.

In pharmacovigilance, most cases are considered only suspected, classified as probable or possible, with rarely definite or unlikely cases.[14] Our study demonstrated this aspect, considering that all 90 cases evaluated by Naranjo algorithm were classified in these categories and 93% of ADR were categorized similarly by WHO-UMC system. The assessment using WHO-UMC is not dependent on a score, allowing a discreet distribution among the categories when compared with Naranjo algorithm. This method has been shown to be easy to assess the causality of ADR, however, the influence of confounding variables was not considered.

Considering preventability, our results revealed about half of cases were considered definitely preventable or probably preventable, in consonance with a meta-analysis conducted by Hakkarainen *et al.* [15] To our knowledge, there is no previous study designed to compare the causality assessment and preventability of ADR developed by multiprofessional teams. We should address some limitations of our study, including fixed sample size, small number of judges and intrinsic interrater variability that hindered to obtain homogeneous results in this investigation. Besides, we did not compare how much better for clinical practice would be the consensus of three professionals from different areas than a consensus of three professionals from the same area.

Finally, to date, a great number of preventable ADR reinforce the need for improvements in pharmacovigilance strategies in hospital settings. Variability in the assessment of ADR could result in heterogeneous decisions in drug safety. Our results suggest that the involvement of interdisciplinary teams could improve the quality of the monitoring process of ADR. Further studies focused on standard definitions and multiprofessional consensual assessment should also be considered as a strategy to seek for improvements in patient safety.

Acknowledgments

This study was supported by the National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Level Education Personnel (CAPES) and the State of Minas Gerais Research Foundation (FAPEMIG), Brazil, and the Pró-reitoria Pesquisa of the Universidade Federal Minas Gerais, Brazil. ALPR is a fellow of CNPq.

Conflict of interest

The authors have no conflicts of interest to declare.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Contributors

R.R.M., M.D.G.S., A.L.P.R., M.A.P.M. participated in the conception and design of the study; M.M.P.F., G.H.M., L.E.C.F. acted as judges of the study; R.R.M. and M.A.P.M. planned and conducted statistical analysis. R.R.M., M.D.G.S., A.L.P.R. and M.A.P.M. participated in the data analysis and interpretation of results. All authors reviewed and approved the final version of the manuscript.

References

1. 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.
2. Thaker SJ, Sinha RS, Gogtay NJ, Thatte UM. Evaluation of inter-rater agreement between three causality assessment methods used in pharmacovigilance. *J Pharmacol Pharmacother* 2016; 7: 31-3.
3. Varallo FR, Planeta CS, Herdeiro MT, Mastroianni PC. Imputation of adverse drug reactions: Causality assessment in hospitals. *PLoS One* 2017; 12: e0171470.
4. Khan LM, Al-Harathi SE, Osman AM, Sattar MA, Ali AS. Dilemmas of the causality assessment tools in the diagnosis of adverse drug reactions. *Saudi Pharm J* 2016; 24: 485-93.
5. 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.
6. Schumock GT, Thornton JP. Focusing on the preventability of adverse drug reactions. *Hosp Pharm* 1992; 27: 538.
7. Shajahan J, Aslam Parathoduvil A, Purushothaman S. An analysis of seriousness, predictability and preventability of adverse drug reactions reported at a tertiary care teaching hospital in Kerala, India: a retrospective observational record based study. *International Journal of Basic & Clinical Pharmacology* 2018; 7: 2433-38.
8. WHO. The Conceptual Framework for the International Classification for Patient Safety. In: World Health Organization, 2009.
9. Silva MDDG, Martins MAP, Viana LG, Passaglia LG, de Menezes RR, Oliveira JAQ, da Silva JLP, Ribeiro ALP. Evaluation of accuracy of IHI Trigger Tool in identifying adverse drug events: a prospective observational study. *Br J Clin Pharmacol* 2018; 84: 2252-59.
10. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971; 76: 378-82.
11. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213-20.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
13. Mouton JP, Mehta U, Rossiter DP, Maartens G, Cohen K. Interrater agreement of two adverse drug reaction causality assessment methods: A randomised comparison of the Liverpool Adverse Drug Reaction Causality Assessment Tool and the World Health Organization-Uppsala Monitoring Centre system. *PLoS One* 2017; 12: e0172830.

14. Center UM. The use of the WHO-UMC system for standardised case causality assessment. https://www.who.int/medicines/areas/quality_safety/safety_efficacy/WHOcausality_assessment.pdf. Accessed Jun 10, 2019

15. Hakkarainen KM, Hedna K, Petzold M, Hagg S. Percentage of patients with preventable adverse drug reactions and preventability of adverse drug reactions—a meta-analysis. PLoS One 2012; 7: e33236.

Table 1. Descriptive data and agreement analyses of causality assessment of ADR	Table 1. Descriptive data and agreement analyses of causality assessment of ADR
Characteristics	Value
ADR duration (<i>days</i>) [mean, (SD)]	3.0 (2.9)
ADR [n, (%)]	
Hypotension	19 (21.1)
Constipation	17 (18.9)
Bleeding	11 (12.2)
Hyperglycemia	10 (11.1)
Renal Injury	10 (11.1)
Somnolence	6 (6.7)
Others ADR ^a	17 (18.9)
Main suspect drugs [n, (%)]	
Morphine	10 (11.1)
Tramadol	8 (8.9)
Captopril	5 (5.6)
Clonazepam	5 (5.6)
Enoxaparin	5 (5.6)
Furosemide	5 (5.6)
Warfarin	5 (5.6)
Carvedilol	3 (3.3)
Enalapril	3 (3.3)
Prednisone	3 (3.3)
Vancomycin	3 (3.3)
Others drugs ^b	35 (38.8)
Naranjo algorithm categories [n, (%)]	
Certain	0 (0)
Probable	29 (32.2)
Possible	61 (67.8)
Doubtful	0 (0)
WHO-UMC system categories [n, (%)]	
Certain	1 (1.1)
Probable	37 (41.1)
Possible	47 (52.2)
Unlikely/Conditional	5 (5.6)
Interrater agreement	
Exactly agreement proportion Judge A x Judge	0.53
Judge B x Judge C	0.24
Judge A x Judge C	0.33
Multiple Judges	0.19
Extreme disagreement proportion	
Judge A x Judge B	0.19
Judge B x Judge C	0.33
Judge A x Judge C	0.31

Table 1. Descriptive data and agreement analyses of causality assessment of ADR

Multiple Judges
Kappa^c; CI (95%)
Judge A x Judge B
Judge B x Judge C
Judge A x Judge C
Multiple Judges
ADR: Adverse drug reaction; SD: Standard deviation; CI: Confidence interval ^a Others ADR included: nausea/vomiting, diarrhea, changing in respiratory pattern, hypoglycemia, rash and tachycardia. ^b Others drugs included 19 drugs with frequency of use [?] ² : Atenolol, bisacodyl, codein/acetaminophen, ciclosporin, insulin, sinvastatin, amiodarone, cefepime, dexamethasone, dobutamine, phenytoin, fenoterol, heparin, hydralazine, gentamicin, hydrocortisone, methylprednisolone, mycophenolate mofetil, rifampicin/isoniazid/pyrazinamide/ethambutol. ^c Cohen's kappa was used for pairwise agreement and Fleiss' kappa for multiple judges.

Table 1. Descriptive data and agreement analyses of causality assessment of ADR

0.29
0.180 (0.04-0.32)
0.168 (0.05-0.29)
0.113 (0.01- 0.22)
0.096 (0.01-0.18)
ADR: Adverse drug reaction; SD: Standard deviation; CI: Confidence interval ^a Others ADR included: nausea/vomiting, diarrhea, changing in respiratory pattern, hypoglycemia, rash and tachycardia. ^b Others drugs included 19 drugs with frequency of use [?] ² : Atenolol, bisacodyl, codein/acetaminophen, ciclosporin, insulin, sinvastatin, amiodarone, cefepime, dexamethasone, dobutamine, phenytoin, fenoterol, heparin, hydralazine, gentamicin, hydrocortisone, methylprednisolone, mycophenolate mofetil, rifampicin/isoniazid/pyrazinamide/ethambutol. ^c Cohen's kappa was used for pairwise agreement and Fleiss' kappa for multiple judges.

Hosted file

Table 1.docx available at <https://authorea.com/users/317506/articles/447582-causality-assessment-of-adverse-drug-reactions-detected-in-a-brazilian-public-hospital-interrater-agreement-using-a-global-introspection-method>