


BMJ Open Quality Validation of a method to assess the severity of medication administration errors in Brazil

Lindemberg Assunção-Costa ^{1,2}, Charleston Ribeiro Pinto,¹ Juliana Ferreira Fernandes Machado,² Cleidenete Gomes Valli,³ Luis Eugenio Portela Fernandes de Souza,⁴ Bryony Dean Franklin⁵

To cite: Assunção-Costa L, Pinto CR, Machado JFF, *et al.* Validation of a method to assess the severity of medication administration errors in Brazil. *BMJ Open Quality* 2023;**12**:e002510. doi:10.1136/bmjopen-2023-002510

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-002510>).

Received 21 July 2023

Accepted 3 December 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Faculty of Pharmacy, Federal University, Salvador, Brazil

²INAF, National Institute of Pharmaceutical Assistance and Pharmacoeconomics, Salvador, Brazil

³Bahia State Health Secretariat, Salvador, Brazil

⁴Federal University of Bahia Institute of Collective Health, Salvador, Brazil

⁵University of London School of Pharmacy, London, UK

Correspondence to
Professor Lindemberg Assunção-Costa;
lindemb@ufba.br

ABSTRACT

Introduction Medication errors are frequent and have high economic and social impacts; however, some medication errors are more likely to result in harm than others. Therefore, it is critical to determine their severity. Various tools exist to measure and classify the harm associated with medication errors; although, few have been validated internationally.

Methods We validated an existing method for assessing the potential severity of medication administration errors (MAEs) in Brazil. Thirty healthcare professionals (doctors, nurses and pharmacists) from Brazil were invited to score 50 cases of MAEs as in the original UK study, regarding their potential harm to the patient, on a scale from 0 to 10. Sixteen cases with known harmful outcomes were included to assess the validity of the scoring. To assess test–retest reliability, 10 cases (of the 50) were scored twice. Potential sources of variability in scoring were evaluated, including the occasion on which the scores were given, the scorers, their profession and the interactions among these variables. Data were analysed using generalisability theory. A G coefficient of 0.8 or more was considered reliable, and a Bland-Altman analysis was used to assess test–retest reliability.

Results To obtain a generalisability coefficient of 0.8, a minimum of three judges would need to score each case with their mean score used as an indicator of severity. The method also appeared to be valid, as the judges' assessments were largely in line with the outcomes of the 16 cases with known outcomes. The Bland-Altman analysis showed that the distribution was homogeneous above and below the mean difference for doctors, pharmacists and nurses.

Conclusion The results of this study demonstrate the reliability and validity of an existing method of scoring the severity of MAEs for use in the Brazilian health system.

INTRODUCTION

Assessing the severity of medication errors is crucial for improving patient safety during medication use. Such an assessment makes it possible to differentiate errors in relation to their severity and thus establish risk minimisation strategies targeting errors with the greatest potential for harm.¹

The assessment of potential and actual harm involves different processes, each

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ In addition to estimating the rate of medication administration errors, it is essential to understand their potential severity in order to propose harm mitigation measures.
- ⇒ There are several existing scales to assess the potential severity of medication administration errors; however, only two have been internationally validated, and only in the context of developed countries.

WHAT THIS STUDY ADDS

- ⇒ We found an existing method of assessing the potential severity of medication administration errors to be valid and reliable for use in Brazil. Severity scores were generally higher than those given by judges from the UK and Germany in previous studies.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ We have validated an existing method for assessing the potential severity of medication administration errors for use in the Brazilian context. Therefore, it could be used in future research, contributing to a better understanding of the severity of medication administration errors.

including two steps: (1) identifying the potential or actual harm to the patient related to a medication error and (2) rating the degree or severity of that harm.² Various tools exist to measure and classify the harms associated with medication errors. For example, a systematic review of harm related to prescription errors identified over 40 harm classification tools used before 2013.³ Among these, the The National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)⁴ and Dean and Barber⁵ methods have been validated internationally. A strength of the latter method is that it can be used to assess the potential severity of medication errors that do not have a known outcome, unlike the NCC MERP, in which the patient outcome must be known.⁵

The Dean and Barber scale was developed to assess the potential severity of medication administration errors (MAEs) by calculating the mean scores of four healthcare professionals (including at least one pharmacist, one nurse and one physician). This method has already been used to assess the potential clinical significance of MAEs identified in studies conducted in the UK and Germany, and is valid and reliable in the contexts in which these studies were conducted.^{5,6}

A recent systematic review of MAEs detected by the direct observation method in Latin American hospitals identified 10 studies that estimated the rate of MAEs; however, none of them assessed the severity of these errors.⁷ Considering the differences between Brazil and countries such as Germany and the USA regarding health systems, professional training and performance, and cultural contexts, it is necessary to validate the method within the Brazilian context.

Therefore, this study aimed to validate the existing Dean and Barber method,⁵ originally developed in the UK, for assessing the potential clinical significance of MAEs in Brazil. Specific objectives were to (a) determine the minimum number of judges required for a reliable mean severity score; (b) determine the effect of a judge's profession on the score and (c) test the validity of the mean score in relation to cases with known clinical outcomes, all in the Brazilian context.

METHODS

The present study

We used the method developed by Dean and Barber to assess the potential severity of MAEs.⁵ The method was described in detail in a published protocol.⁸

In brief, 37 healthcare professionals (doctors, nurses and pharmacists) from Brazil were invited to participate in scoring 50 cases of MAEs gathered from an original UK study⁵ regarding their potential harm to patients on a scale of 0–10. By looking at 10 cases (out of 50) scored twice, reliability was assessed, and potential sources of variability in scoring were assessed depending on the error case, the occasion when the scores were given, the scorer, their profession and the interactions among these variables. We chose to have just 10 cases assessed twice to avoid excessive workload for the judges and reduce the risk of non-response. This number was also sufficient in the earlier study.⁵

Generalisability theory was used to analyse the data.

The previous methodology

When creating their method, Dean and Barber chose 50 medication error cases from the literature in nearly equal numbers showing minor, moderate and severe potential clinical outcomes; in 16 of these cases, the patient outcome was already known. These cases were then sent to 30 different healthcare professionals (10 physicians, 10 nurses and 10 pharmacists). These judges were asked to score the potential clinical significance

on a Visual Analogue Scale ranging from 0 to 10 (with 0 corresponding to 'no harm' and 10 corresponding to 'patient death'). Specifically, this error severity classification involves: (1) Minor—very unlikely that the patient will develop any adverse event; (2) Moderate—likely to cause an adverse event in the patient or interfere with the therapeutic goal, but very unlikely to cause death or harm lasting more than a week and (3) Serious—an error that could lead to permanent harm or death to the patient. A subset of 10 cases was evaluated on a second occasion by all judges. The data were analysed using generalisability theory.

Judge recruitment

Judges were recruited from eight large general hospitals. Hospitals were chosen to give geographical diversity, since Brazil is a country of continental dimensions with potential for regional differences in practice. Four of the five regions of Brazil were represented.

The heads of service at these hospitals were contacted to identify doctors (internists, generalists and clinical specialists), nurses and pharmacists with more than 3 years of experience willing to evaluate the potential severity of the 50 medication errors.

After the heads of service's acceptance, 37 professionals were contacted, invited by email and the consent forms and letters describing the scoring process guidelines, the objectives of the study and practical examples of how to perform the scoring were sent to them, regardless of the location. No incentive was offered to professionals to participate in this study.

Scoring process

The invited professionals received a file with the descriptions of 50 cases of EAMs and were instructed to score the cases according to their potential clinical significance, using the scale proposed by Dean and Barber. The scores provided by these professionals were then analysed. Two weeks after receiving the severity ratings based on the 50 cases, each respondent received a random sample of 10 of the 50 cases for rescore; selected using the RV.BINOM function within SPSS 29.0.

That way, it was possible to measure whether the occasion on which the cases were scored was an important source of variance in the responses obtained.

Raters were instructed to record the time spent evaluating all 50 cases and to make relevant comments about the scoring process in a specific space on the form, in addition to completing a brief questionnaire about demographic details, including their occupation and number of years of professional experience, along with any comments about the scoring process).

Translation and contextualisation of the cases

The 50 cases were translated by the principal investigator into Portuguese, updated (if the drugs were no longer available or not in routine use), and adapted to the Brazilian context (making necessary adjustments

regarding drugs, doses, concentrations, units of measurement, pharmaceutical forms and available presentations to ensure that all were medications routinely used in clinical practice in Brazilian hospitals) (online supplemental appendix A).

The translated and adapted versions were double-checked by two experienced hospital pharmacists to ensure that the meaning of each case of MAEs remained unchanged. We used the same cases performed in the UK and Germany^{5,6} to allow comparison with the current study.

Generalisability theory

Cronbach *et al*⁹ developed generalisability theory, a method that systematically allows the effect of multiple sources of variance and their interactions on scores to be measured simultaneously in a single study, based on the premise that in any assessment procedure, variance in scores can be attributed to different identifiable sources.

Generalisability theory also emphasises the estimation of variance components. Once the variance attributed to each source is calculated, the most efficient method for reducing unwanted variations can be determined. The results can be used to identify methods for improving the reliability of a test.¹⁰

The application of generalisability theory occurs in stages. First, generalisability analysis begins with the specification of a universe of admissible observations through the identification of different sources of variation. In the second stage, the generalisability study ('G-study'), estimates the variance components of this universe. This involves creating an appropriate research design, collecting data and determining the extent to which each variable influences scores. Different coefficients of variation are calculated to represent the different situations. For example, a coefficient can be calculated to show the extent to which the score assigned to a case by a physician can be generalised to that assigned to the same case by a pharmacist. The final step is a decision study ('D-study') associated with a prespecified universe of generalisability.^{9,10} Broadly speaking, D-studies emphasise the estimation, use and interpretation of variance components for decision-making, with well-specified measurement procedures.¹¹ Perhaps the most important D-study to consider is the specification of a universe of generalisation, where the universe to which a decision-maker wants to generalise is based on the results of a D-study using a particular measurement procedure.¹⁰ From the estimated variance, the effect of a change in the number of observations on the generalisability coefficient can be explored.

Reliability analysis

Universe of observations

We used the same approach to analysis as that used previously.⁵ The sources of variance in the process of assessing the MAEs errors were those inherent in the cases themselves ('case'), the occasion on which they are assessed ('occasion'), the evaluator ('judge'), the professional

background of the judge ('profession') and the interactions among these. Since each judge is a member of a single profession, the judge factor was considered to be nested within the profession factor ('judge: profession').

Because the scores for the 50 cases of errors were obtained on two occasions for a sample of 10 cases, there are two models for conducting the G-study, depending on the data set used:

Model 1: occasion \times case \times judge (using the 10 cases scored twice).

Model 2: case \times judge: profession (using all 50 cases).

Models 1 and 2 ignored the effect of profession and occasion, respectively. A model that considers all sources of variance for the 10 cases with repeated scores, occasion \times case \times judge: profession, was not used because the variance per case was anticipated to be too high to perform an analysis of variance (ANOVA).

Generalisability study

The data were evaluated using models 1 and 2 to determine the contribution of each factor to the variance in scores. First, repeated measures of variance analysis was performed and seven sources of variance estimated for model 1: case, occasion, occasion \times case, judge, judge \times case, occasion \times judge and judge \times case \times occasion. For model 2, the sources of variance were profession, judge 'nested' in profession, case, case \times profession and residual variance (case \times judge: profession). Online supplemental appendix B provides the equations used to calculate the generalisability coefficients.

The resulting mean square values were then used to calculate the attributable variance for each source, using equations for the mean squares based on those described by Streiner and Norman¹² and Cronbach *et al*.⁹ When the estimated variance components were computed as negatives, a value of 0 was assumed.¹³ The overall generalisability coefficient, coefficients equivalent to inter-rater reliability and test-retest reliability were computed.

Decision study

The D-study was based on the G-study results and obtained the necessary decision-making information for the reliable use of generalised scoring scales. D-study was also used with model 1 to calculate G coefficients that identified the required number of judges to achieve sufficient reliability of the scale's usage.

In the D-study, the effects of different modifications in the evaluation procedure on the generalisability coefficient were investigated, and the accuracy of the obtained measurement results evaluated. Therefore, different scenarios based on the results of the G-study were investigated. The same model as in the G-study was used to calculate the generalisability coefficients for different numbers of judges and occupations. This allowed for identifying the number of judges needed to obtain a reliable average score. The D-study also investigated whether judges needed to be from a different or similar profession. Generalisability coefficients for different numbers



of judges and test occasions were calculated using the formula described by Streiner and Norman.¹² As in previous studies, a generalisability coefficient greater than 0.8 was taken to represent acceptable reliability.⁵

Validity analysis

Sixteen cases (out of 50) with known harmful outcomes were included to assess the validity of the scoring process. These cases were the same as those described in Dean and Barber's assessment method: five MAEs reported in the literature that resulted in minor outcomes (no noticeable adverse effects), five with moderate outcomes (some adverse effects but no lasting impairment) and six with severe outcomes (death or lasting impairment). The mean scores assigned by the 30 judges to these 16 cases were examined to assess whether they reflected the severity. Thus, it was possible to test the method's validity by comparing the scores assigned by the 30 raters to the 16 MAEs with previously established scores.

Test–retest reliability

The test–retest agreement was assessed by performing the Bland-Altman test, as an enhancement to the earlier studies. The Bland-Altman plot helps visualise and interpret the test–retest agreement. By definition, 95% of the differences between repeated measures must be within

agreed-upon limits. We used the Bland-Altman method as an alternative way to test the reliability.

Comparison between the analyses performed in Germany, the UK and Brazil

The mean scores of the studies in Brazil were first compared with those in the UK⁵ and second with those in the Germany⁶ studies. A paired sample t-test was performed, with a significance level set at 95% CI.

Analyses

All analyses were conducted using the R programming language, V.4.0.3 and SPSS 29.0

RESULTS

Judge recruitment

The heads of service at 8 Brazilian hospitals were contacted to identify doctors, nurses and pharmacists willing to evaluate the potential severity of the 50 medication errors. One of the south-eastern hospitals was unable to participate owing to time constraints. The seven participating hospitals were located in four regions of the country (three in the southeast, two in the northeast, one

Table 1 Initially contacted professionals, response rate and final participants

Hospital	Region	Professional group	Letters sent	Response	Response rate (%)	No of final participants
HUPES	NE	Doctors	6	4	67	4
HCPA	SO	Doctors	2	2	100	2
HCUFMG	SE	Doctors	2	2	100	2
HB	NE	Doctors	1	1	100	1
HGV	NO	Doctors	1	1	100	1
HUPES	NE	Nurses	2	2	100	2
HCPA	SO	Nurses	2	2	100	2
HCUFMG	SE	Nurses	2	1	50	1
HB	NE	Nurses	2	2	100	2
HSL	SE	Nurses	2	2	100	2
HGV	NO	Nurses	1	1	100	1
HUPES	NE	Pharmacists	5	3	60	2
HCPA	SO	Pharmacists	2	2	100	2
HCUFMG	SE	Pharmacists	2	2	100	2
HB	NE	Pharmacists	2	2	100	2
HSL	SE	Pharmacists	1	0	0	0
João XXIII	SE	Pharmacists	2	1	50	1
HGV	NO	Pharmacists	1	1	100	1
Total			37	31	84	30

HB, Hospital da Bahia; HCPA, Hospital das Clínicas de Porto Alegre; HCUFMG, Hospital das Clínicas da Universidade Federal de Minas Gerais; HGV, Hospital Getúlio Vargas; HSL, Hospital Sírio Libanês; HUPES, Hospital Universitário Professor Edgard Santos; João XXIII, Hospital João XXIII; NE, Northeast; NO, North; SE, Southeast; SO, Southwest.

Table 2 Analysis of variance

Sources of Variation	df	Sum of squares	Mean sum of squares
Case	9	3224	358.2
Judge	29	496	17.1
Occasion	1	2	2.3
CasexJudge	261	990	3.8
CasexOccasion	9	16	1.7
JudgexOccasion	29	665	22.9
CasexJudgexOccasion	261	932	3.6

*Source: Assunção-Costa (2022)

in the south and one in the north), six of whom declined participation. The first 30 professionals who completed the questionnaire were selected for the study. Ultimately, 30 healthcare professionals participated in the study, including 10 nurses, 10 pharmacists and 10 doctors, who responded to the same protocol in two different instances, as shown in [table 1](#).

All 30 judges submitted completed forms for all 50 cases including 10 repeated cases of MAE evaluations, with the absence of one judge who completed the scoring but did not report the time taken. The time taken for each of the remaining 29 judges to score all 50 cases ranged from 14 to 53 min, with a mean of 26.3 min. The mean score for each MAE ranged from 1.6 to 9.3 (online supplemental appendix B).

Two judges (one doctor and one pharmacist) commented on the scoring process and case clarification.

Example 1: 'I faced doubts regarding certain questions, including the lack of knowledge about two to three medications' further serious adverse events'.

Example 2: 'I found it difficult to contextualise the case-specific available information and to separate process error analysis from the analysis of the patient's potential harm'.

Generalisability study

Model 1

[Table 2](#) presents the ANOVA for model 1 (occasion×case×judge). professionals, including doctors, pharmacists and nurses, who separately evaluated 10 identical MAE cases on 2 separate occasions, participated in the analysis.

Online supplemental appendix C presents the estimated variance components that portray the main source of variance as the difference between MAE cases, followed by the 'judge×occasion' design. Occasion was not an important source of variance. The overall generalisability coefficient was 0.99.

Online supplemental appendices D–F presents a constant G coefficient among the three judges, regardless of profession, given by doctors, pharmacists and nurses, respectively.

Table 3 Sources of variance (doctors, pharmacists and nurses)

Sources of variation	df	Sum of squares	Mean sum of squares
Case	49	6068	123.84
Profession	2	577	288.42
Profession×judge	27	1857	68.77
Casexprofession	98	585	5.97
Casexprofession×judge	1323	4653	3.52

Source: Assunção-Costa, 2022.

Model 2

This model was evaluated by all 30 participants simultaneously, using all 50 cases. [Table 3](#) presents the sources of the variance results. [Table 4](#) presents the number needed to obtain reliable G-coefficients, and [figure 1](#) presents a graph of these estimates, considering the different professions of doctors, pharmacists and nurses. Online supplemental appendix G presents the G-studies and D-studies of doctors, pharmacists and nurses.

The G coefficients were calculated considering judges from different professions and are presented in online supplemental appendix H. For example, a pharmacist, nurse and doctor scoring the same case resulted in a good G coefficient of 0.89.

Validity

[Figure 2](#) shows the mean scores of the 16 cases of known severity. A relationship appears to be existed between the known severity categories and the mean

Table 4 G coefficient estimates to maximise scale reliability for future studies (doctors, pharmacists and nurses)

No of judges	G coefficient
	ρ
1	0.76
2	0.85*
3	0.89
4	0.91
5	0.92
6	0.93
9	0.95
12	0.96
15	0.96
18	0.96
21	0.97

Source: Assunção-Costa, 2022.

(ρ): letter phi has more contributions to the variance. It is a more conservative coefficient representing, therefore, how generalisable the instrument is to different samples in other contexts.

*A generalisability coefficient greater than 0.8 was considered to represent acceptable reliability.

G Coefficient Estimates

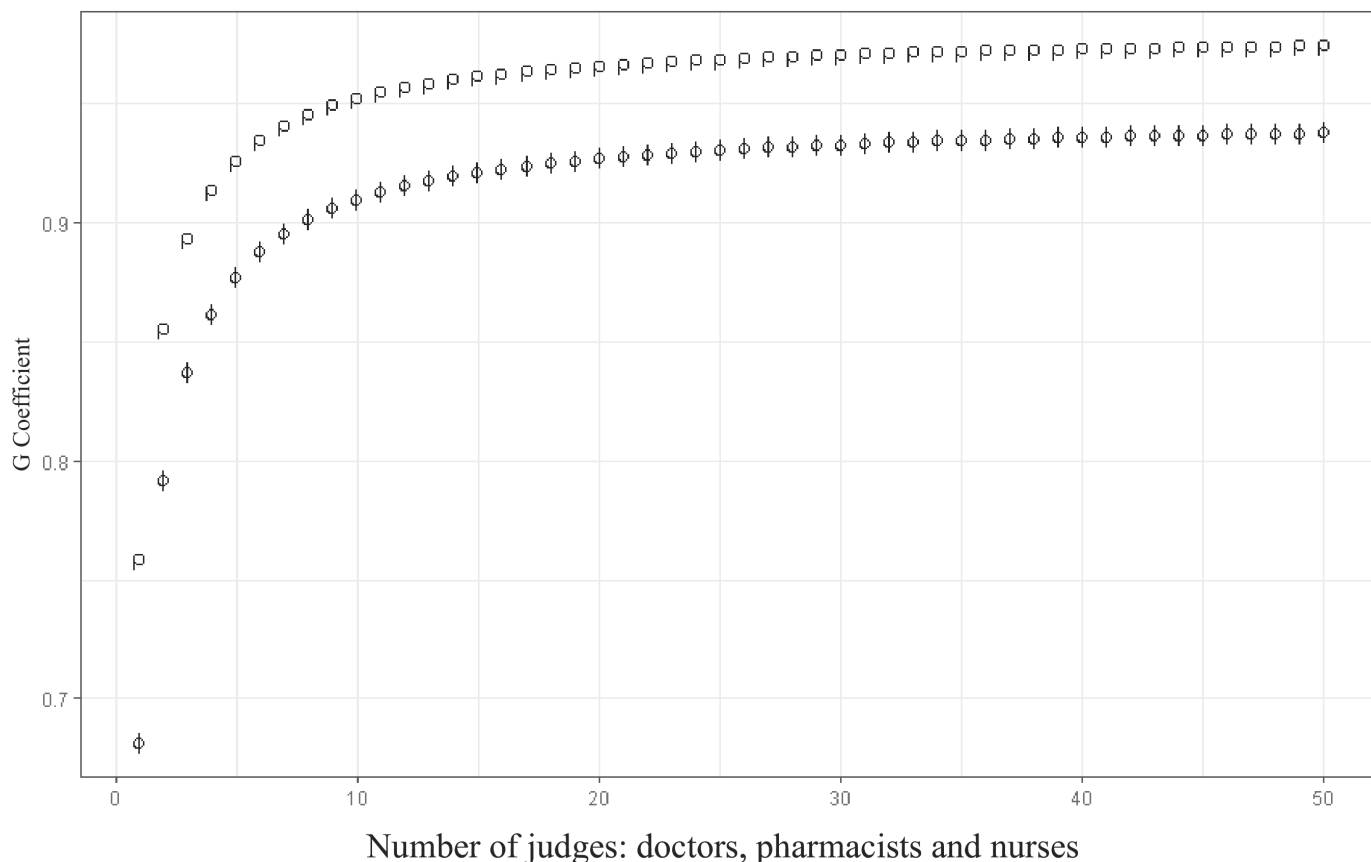


Figure 1 G coefficient estimates to maximise protocol reliability of future studies (doctors, pharmacists and nurses). Embedded text: G coefficients/G coefficients estimates/number of judges—doctors, pharmacists and nurses.

scores assigned by the judges. Minor, moderate and severe cases had mean scores ranging from 2.1 to 5.1, 4.5 to 7.9 and 6.3 to 9.3, respectively. For two minor severity cases (cases 5 and 22), the calculated mean scores were compatible with moderate severity. For one moderate severity case (case 15), the calculated mean score was compatible with a severe case. For one severe case (case 22), the calculated mean score was compatible with moderate severity. The judge-based individual scores assigned to each error indicate the contribution of extremely high or low values to these results. The cases with overlapping scores are described in online supplemental appendix I. In general, the mean scores obtained in Brazil were higher than those obtained in the UK⁵ and German⁶ studies (figure 3). The mean scores of Brazilian judges were 1.36 times (95% CI 1.11 to 1.62; $p < 0.001$; paired samples t-test $t(49) = 10.669$) and 0.49 times (95% CI 0.24 to 0.74; $p < 0.001$; paired samples t-test $t(48) = 4.046$) higher than those of the UK⁵ and German judges,⁶ respectively.

Regarding the case of number 22, which involves the drug paracetamol, there was a difference of 2.6 more points in the average score given by the Brazilian judges in relation to the known scores.

Only six of the four cases had a mean score lower than the scores calculated for the UK and German judges.^{5 6} The maximum difference between scores obtained in Brazil and the UK was 2.7, and 3.9 between Brazil and Germany.⁶

Test-retest reliability

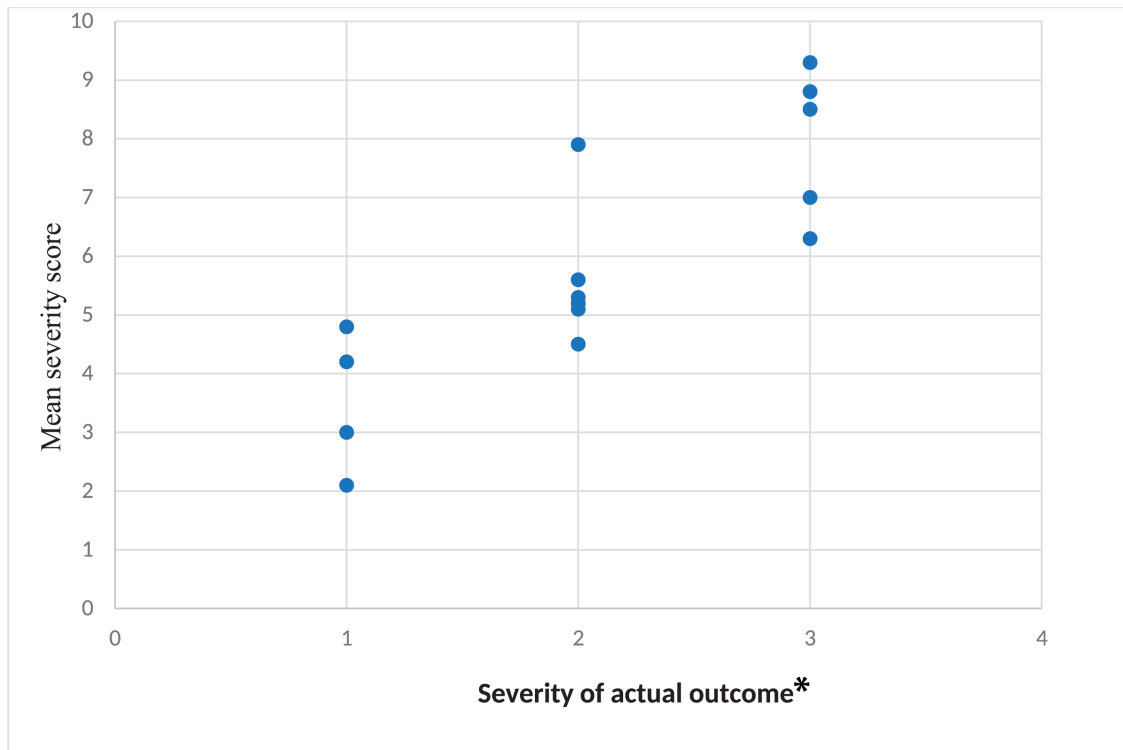
The distribution was homogeneous above and below the mean difference between the two instances with $p = 0.96$ for the doctor, 0.63 for the nurses and 0.38 for the pharmacists. The Bland-Altman plots for doctors, nurses and pharmacists are illustrated in online supplemental appendix J.

DISCUSSION

Key findings

Our findings indicate the suitability of Dean and Barber's⁵ MAE clinical severity scoring scale for use in the Brazilian healthcare system. A Brazilian doctor, nurse and pharmacist's mean score results are reliable and valid, owing to their potential generalisation to the same group of health professionals and because they allow differentiation of minor, moderate and severe errors.

The reliability of this method in Brazil had a remarkable resemblance to the original British study as well



* 1 = minor, 2 = moderate, 3 = severe

Figure 2 Comparing the judges' mean scores and the actual outcome severity. *1=minor, 2=moderate, 3=severe.

as the German research by Taxis *et al*⁶ based on similar coefficients. In both studies, the variance was not significantly affected by judges or their professions. Compared with German and English judges, Brazilian judges also took similar lengths of time to score the cases. We reported results similar to those of Taxis

*et al*⁶ in their German study, which concluded that three judges from different professions were sufficient to obtain a reliable mean score, in contrast to the requirement for four judges reported in the English study. Taxis *et al*⁶ claim the origin of this difference to be the model used in the D-study that calculated the

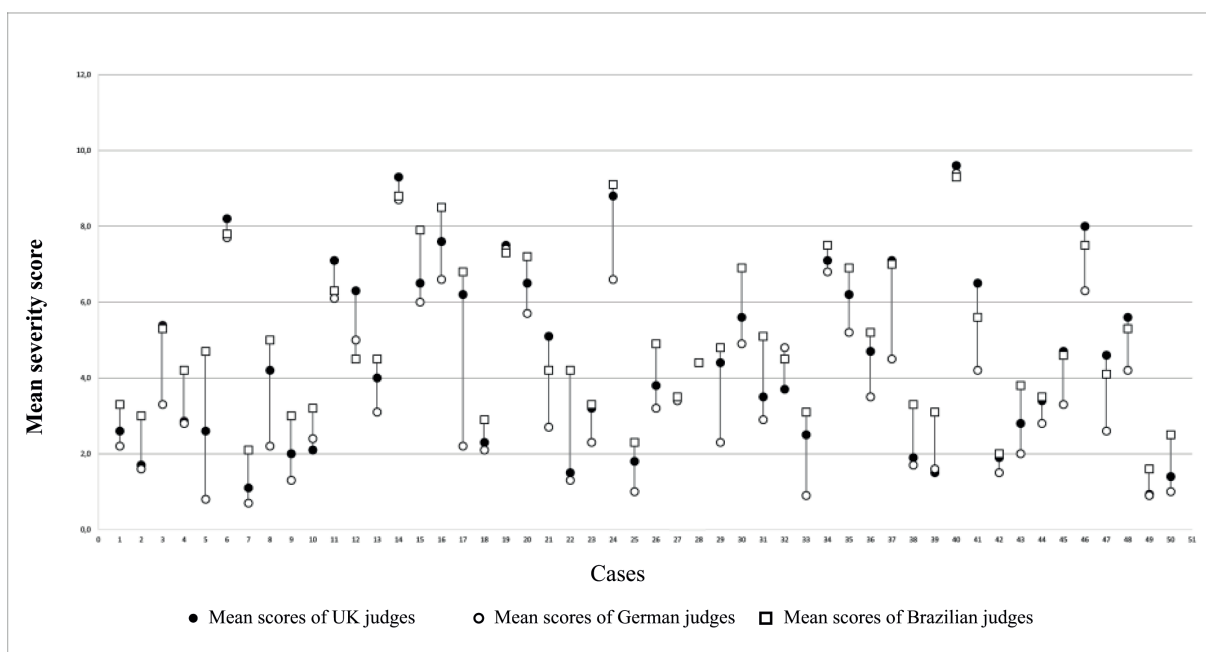


Figure 3 Mean score comparison of Brazil, Germany and the UK.

generalisation coefficient through the ‘no occasion’ facet, which contributed minutely to the variance. In our study, occasion was not an important source of variance.¹⁴

A novel aspect of our study lies in the additional evidence of reliability obtained using the Bland-Altman analysis, which also confirmed the agreement between the responses of each sampled professional provided at two separate instances, corroborating the results obtained by generalisability theory. We used the Bland-Altman method as an alternative way to test the reliability.¹⁴

Implications for practice and future research

This study confirms that this method can also be used in Brazil to assess the severity of medication errors and that the scale is valid for differentiating between MAES with minor, moderate and severe outcomes. There is a debate regarding the instruments for assessing the severity of medication errors and their ability to reflect the harmful effects on patients. This is either due to the absence of an ideal assessment method for the scale’s validity or cases of the validation process not reflecting actual regular cases, thus leading to interpretation biases.⁶ Newly developed tools reduce uncertainties in this evaluation, yet they so far lack validation.¹⁵

In general, judges considered errors with a mean score of less than two as minor errors, posing a low probability of harm to patients. In contrast, mean scores above two, considered moderate and severe, can be attributed to errors that adversely affect patients. Our results corroborated those of Taxis, Dean and Barber.⁶

Overall, the scores were higher in the present study than in the original one. For example, there was a mean score overlap in two minor cases with mean scores assigned to moderate or severe errors. In general, judges in Brazil considered errors more serious than previous judges in the UK and Germany. For example, Brazilian judges considered it more serious for an elderly patient to have a double dose of paracetamol, (case 22), compared with scorers in the UK⁵ and Germany.⁶ It is not clear to what extent this reflects international differences in safety culture or clinical practice, or changes over time in understanding or attitudes to risk. Further research is needed to explore these issues.¹⁶ For example, a contemporaneous study comparing scores among countries would establish whether differences can be attributed to countries rather than reflecting changes over time.

Strengths

This evaluation process included physicians, pharmacists and nurses with experience in the clinical field and working in hospitals distributed across four regions of the country, which provided an overview of the assessment of the potential severity of medication errors in Brazil.

Limitations

A potential limitation is the use of cases from an earlier UK study rather than selecting MAEs from the Brazilian

context. However, using the same or similar cases allowed us to make a more meaningful comparison among countries. There is also very little literature about MAEs in Brazil from which to select suitable cases. The medications described in the errors were all in common use in Brazilian hospitals at the time other studies.

Judges were not chosen at random but were required to have at least 3 years of clinical practice and represented a range of public and private hospitals from different geographical areas. Depending on the judges’ selection criteria, one institution may have contributed the majority of professionals. Such was the case of the northeastern region, which was a potential limitation.

CONCLUSION

The results of this study demonstrate the validity and reliability of Dean and Barber’s scale for assessing the severity of MAEs in the Brazilian health system.

Twitter Lindemberg Assunção-Costa @lindembergrn and Bryony Dean Franklin @BryonyDF

Acknowledgements We acknowledge and thank Wagner Franco for translating the article as well as Flavia Del Castanhel and Leonardo Maia Nogueira for their collaboration in the statistical analysis and intellectual support with G Theory. We would also like to thank the 30 healthcare professionals who collaborated in assessing the severity of the 50 cases of medication errors.

Contributors LA-C and LEPFdS made substantial contributions to the conception or design of the work. Material preparation, data collection and analysis were performed by LA-C, CGV, CRP and JFFM. The first draft of the manuscript was written by LA-C, JFFM, BDF and CGV. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. This work is based primarily on the doctoral thesis. Acting as guarantor LA-C.

Funding This work had the support of the National Institute of Pharmaceutical Assistance and Pharmacoeconomics (INAFF). INAFF is a national Brazilian non-governmental organization dedicated to promote access and rational use of medicine.

Disclaimer The views expressed are those of the authors and not necessarily those of INAFF, NHS, NIHR or the Department of Health and Social Care

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and this study was submitted to the evaluation of the Research Ethics Committee of the Complexo Hospitalar Universitário Professor Edgard Santos and approved under opinion number 3.102.570/2019. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information. Not applicable.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Lindemberg Assunção-Costa <http://orcid.org/0000-0002-8752-7301>

REFERENCES

- 1 Walsh EK, Hansen CR, Sahm LJ, *et al*. Economic impact of medication error: a systematic review. *Pharmacoepidemiol Drug Saf* 2017;26:481–97.
- 2 Morimoto T, Gandhi TK, Seger AC, *et al*. Adverse drug events and medication errors: detection and classification methods. *Qual Saf Health Care* 2004;13:306–14.
- 3 Garfield S, Reynolds M, Dermont L, *et al*. Measuring the severity of prescribing errors: a systematic review. *Drug Saf* 2013;36:1151–7.
- 4 National Coordinating Council for Medication Error Reporting and Prevention. NCCMERP index for categorizing medication errors, Available: <https://www.nccmerp.org/sites/default/files/indexColor2001-06-12.pdf>
- 5 Dean BS, Barber ND. A validated, reliable method of scoring the severity of medication errors. *Am J Health Syst Pharm* 1999;56:57–62.
- 6 Taxis K, Dean B, Barber N. The validation of an existing method of scoring the severity of medication administration errors for use in Germany. *Pharm World Sci* 2002;24:236–9.
- 7 Assunção-Costa L, Costa de Sousa I, Alves de Oliveira MR, *et al*. Drug administration errors in Latin America: a systematic review. *PLoS One* 2022;17:e0272123.
- 8 Assunção-Costa L, Ribeiro Pinto C, Ferreira Fernandes Machado J, *et al*. Validation of a method to assess the severity of medication administration errors in Brazil: a study protocol. *J Public Health Res* 2022;11:2623.
- 9 Cronbach LJ, Gleser GC, Nanda H, *et al*. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley & Sons, 1972.
- 10 Brennan RL. *Generalizability Theory*. New York, NY: Springer, 2001.
- 11 Del Castanhel F, Fonseca FR, Bonnassis Burg L, *et al*. Applying the Generalizability theory to identify the sources of validity evidence for the quality of communication questionnaire. *The American Journal of Hospice & Palliative Care* 2023;10499091231201546.
- 12 Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press, 1995.
- 13 Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. *Am Psychol* 1989;44:922–32.
- 14 Assunção-Costa L. Erros de Administração de Medicamentos em um hospital Universitário: Incidência, Natureza, Gravidade e Fatores Associados. PhD thesis. 2022. Universidade Federal da Bahia, Associados.
- 15 Gates PJ, Baysari MT, Mumford V, *et al*. Standardising the classification of harm associated with medication errors: the harm associated with medication error classification (HAMEC). *Drug Saf* 2019;42:931–9.
- 16 Fahmy S, Garfield S, Furniss D, *et al*. A comparison of two methods of assessing the potential clinical importance of medication errors. *Safety in Health* 2018;4.