

A COMPARISON OF CAUSALITY ASSESSMENT TOOLS FOR SUSPECTED ADVERSE DRUG REACTIONS IN HOSPITALIZED PATIENTS AT A TERTIARY CARE HOSPITAL

PARIKH CD¹, DESAI CK², KAPADIA JD², SHAH MK²

¹Department of Pharmacology, Nootan Medical College and Research Centre, Visnagar, Gujarat, India. ²Department of Pharmacology, B.J. Medical College, Ahmedabad, Gujarat, India. Email: jigar.d.kapadia@gmail.com

Received: 28 April 2020, Revised and Accepted: 08 June 2020

ABSTRACT

Objective: The objective of the study was to compare six causality assessment (CA) tools for suspected adverse drug reactions (ADRs) reported in hospitalized patients at a tertiary care hospital in India.

Methods: Intensive ADR monitoring was performed in indoor patients of two randomly selected medicine units. A detailed case report of each suspected ADR (n=120) was provided to six independent experts for CA using either visual analog scale (VAS) or WHO-UMC scale. Investigator assessed causality using Naranjo's scale, Koh *et al.* scale, the French method, and Karch and Lasagna scale. Similar causality categories from these scales were coded for correlation. Agreement among experts and that between various CA tools were analyzed using Cohen's kappa and Fleiss kappa. Reasons for disagreements among different scales were evaluated.

Results: A variation was observed in the total number of drugs suspected to cause ADR by experts and investigator. "Likely" and "Plausible" causality were suggested frequently by experts using VAS whereas "Possible" causal association was frequent according to experts using the WHO-UMC scale and also by the investigator using algorithms except Koh *et al.* scale. None to the slight agreement was observed among experts who used VAS (k=0.117), whereas a substantial agreement was observed among experts using the WHO-UMC scale (k=0.707). A substantial agreement was observed between Karch and Lasagna scale and the French method (k=0.740). Both scales demonstrated moderate agreement with Naranjo's scale. Disagreement among the WHO-UMC scale, the French method, and Karch and Lasagna scale were associated with polypharmacy, serious ADRs, non-availability of laboratory data, and skin and subcutaneous tissue ADRs.

Conclusion: A higher inter-rater agreement with the WHO-UMC scale suggests its utility for CA of suspected ADRs in indoor patients. The French method and Karch and Lasagna scale can be used for CA in hospitalized patients as an adjunct to Naranjo's scale. Factors associated with disagreement should be considered at the time of reporting ADRs and evaluating causality.

Keywords: Suspected adverse drug reaction, Agreement, Algorithm, Causality assessment, Intensive monitoring, Visual analog scale, WHO-UMC scale.

© 2020 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2020.v13i7.38067>

INTRODUCTION

Causality assessment (CA) is defined as the evaluation of the possibility of a drug being the cause of an adverse drug reaction (ADR). It measures the strength of the relationship between a drug and suspected ADR. It is performed to identify important ADRs, to generate signals, and to evaluate the risk-benefit profile of drugs [1]. A precise and accurate method of CA is key to effective management and minimization of ADRs [2].

Accurate CA is posed with certain challenges such as inadequate data, lack of a universally accepted method, and an inherently complex evaluation. Principle methods of CA include expert judgment, algorithmic approach, and probabilistic method. Expert judgment is widely used and involves an expert who applies knowledge and experience to evaluate causality. However, inter- and intra-rater variation and weak reproducibility are observed with this method [3]. Various scales have been utilized for expert judgment, including WHO-UMC scale [4] and visual analog scale (VAS) [5]. However, a low inter-rater agreement is observed with the WHO-UMC scale as compared to algorithms [6].

Algorithms are simple to use and demonstrate a higher intra and inter-rater agreement. These have poor sensitivity but good specificity as compared to expert judgment and probabilistic methods [3]. A number of algorithms are in use, including Naranjo's scale [7], Koh *et al.* Scale [8], Karch and Lasagna Scale [9], and the French method [10].

Studies conducted previously show poor agreement between the three methods of CA [11] and also among different scales being used [12].

These studies, however, have usually relied on spontaneously reported ADRs, whereas an intensive monitoring of ADRs is likely to provide more robust information for accurate CA. Furthermore, few studies have evaluated the agreement between the WHO-UMC scale, VAS, and the above-mentioned algorithms for ADRs occurring in indoor patients.

With this view, the present study was conducted to evaluate the agreement between above mentioned CA tools for ADRs reported in hospitalized patients and to evaluate the utility of these scales in the Indian health-care system.

METHODS

This was an observational, prospective, and single-center study conducted in indoor patients of two randomly selected units of the Department of Medicine at a Tertiary Care Hospital in Gujarat, India over a period of 23 months, that is, September 2016–August 2018. Permission to conduct the study was obtained from the IEC (Ref. No: IEC/Certi/21/17) and head of the Department of Medicine. Patients hospitalized with an even indoor registration number to the selected medicine units, of either gender and aged more than 12 years, who developed an ADR following admission or who were hospitalized due to an ADR were included after written informed consent. Intensive monitoring of observed ADRs was performed. Enrolled patients were followed up till resolution of ADR or discharge, whichever was earlier. A sample size of 30 ADRs was needed to assess agreement between two scales or methods, as determined from previous studies [13–15]. Hence, a total of 120 ADRs were included in the sample to evaluate the agreement between four algorithms.

Case reports of all ADRs were provided to experts for CA using either the WHO-UMC scale or VAS. Experts 1 (clinician), 2 (clinician), and 5 (pharmacologist) evaluated causality using VAS, whereas experts 3 (clinician), 4 (clinician), and 6 (pharmacovigilance associate) evaluated causality using the WHO-UMC scale. Length of VAS from zero to the mark assigned by the expert was measured by the investigator and converted to causality categories as described by Arimone *et al.* (excluded [0–5 mm], unlikely [6–25 mm], doubtful [26–45 mm], unassessable/ unclassifiable [46–55 mm], Plausible [56–75 mm], likely [76–95 mm], and certain [96–100 mm]) [5]. Drug assigned highest score on VAS or highest category on the WHO-UMC scale was considered as the primary suspect drug by the respective expert. Investigator evaluated causality of the same ADRs using Naranjo's scale, Koh *et al.* scale, the French method, and Karch and Lasagna scale. Severity and preventability of ADRs were assessed by the investigator using the modified Hartwig and Siegel scale [16] and modified Schumock and Thornton criteria [17], respectively.

The causality of ADRs among four algorithms, WHO-UMC scale and VAS was matched using the coding system described by Thaker *et al.* (certain/definite-1, probable/likely-2, possible/plausible-3, doubtful, unassessable/unclassifiable, unlikely, excluded, conditional/unclassified, and unrelated- 4) [6]. Agreement between two scales was calculated using Cohen's kappa [18] and that between > two raters was calculated using Fleiss kappa [19] (k≤0: No agreement, 0.01–0.20: None to slight agreement, 0.21–0.40: Fair agreement, 0.41–0.60: Moderate agreement, 0.61–0.80: Substantial agreement, 0.81–1.0: Almost perfect agreement, and negative value: agreement worse than that expected by chance alone) [20,21].

Factors associated with disagreement among three scales, that is, Karch and Lasagna Scale, the French method, and WHO-UMC scale, were also evaluated using the Chi-square test. For this, ADRs were divided into Group A- ADRs which showed 100% agreement between three scales and Group B- remaining ADRs. $P < 0.05$ was considered as statistically significant.

RESULTS

A total of 2304 hospitalized patients were screened and 120 patients who developed ADRs were enrolled (incidence of ADRs: 5.2%). Enrolled patients suffered from infections (25.8%), gastrointestinal (20.8%), respiratory (18.3%), and cardiovascular (17.5%) disorders. ADRs were frequent (16.7%) in patients of 71–80 years (Fig. 1) as compared to other age groups except those in 13–20 years and 81–90 years.

Observed ADRs included skin and subcutaneous tissue disorders (30%), gastrointestinal system disorders (24.1%), metabolic and nutritional disorders (15.8%), general disorder and administration site conditions (12.5%), neurological disorders (6.7%), liver and biliary system disorders (3.3%), renal and urinary disorders (2.5%), respiratory disorders (1.6%), blood and lymphatic disorders (1.6%), musculoskeletal disorders (0.8%), and investigations (0.8%). Antimicrobials and drugs acting on the renal system were suspected most frequently by investigators and experts. Expert 5 also suspected drugs acting on GI system frequently (Table 1).

CA by Experts using VAS

Experts 1, 2, and 5 suspected a total of 209 drugs for 117 ADRs, 209 drugs for 104 ADRs, and 257 drugs for 120 ADRs, respectively. Experts 1 and 2 assessed three and sixteen cases, respectively, as "not related to drug." Disagreement regarding the primary suspect drug was observed in eight cases. In three cases, all experts suspected different primary drugs. In the remaining 5 cases, experts 1 and 5, experts 2 and 5, and experts 1 and 2 agreed in two, two, and one case, respectively. All experts assigned "Likely" and "Plausible" causality more frequently for the primary suspect drug (65% by expert 1, 77.9% by expert 2 and 74.2% by expert 5). Expert 5 assigned "Certain" causality more frequently (4.2%) than other two experts (2.5% by expert 1 and none by expert 2). For all suspect drugs, "Plausible," "Unassessable," and "Doubtful" associations were suggested frequently by experts. "Unlikely" association was also suggested in fair number of reports (Table 2).

CA by experts using the WHO-UMC scale

Experts 3, 4, and 6 suspected a total of 155, 147, and 149 drugs respectively for 120 ADRs. "Possible" association with the primary suspect drug was reported in the majority of cases (74.1%, 80%, and 75.8% by experts 3, 4, and 6, respectively) (Fig. 2). A "Possible" causal association between other suspect drugs and ADR was assigned in 35 (29.16%), 27 (22.5%), and 29 cases (24.16%) by experts 3, 4, and 6, respectively.

CA by investigator using algorithms

The investigator suspected a total of 190 drugs for 120 ADRs. "Possible" association with the primary suspect drug was most frequent using

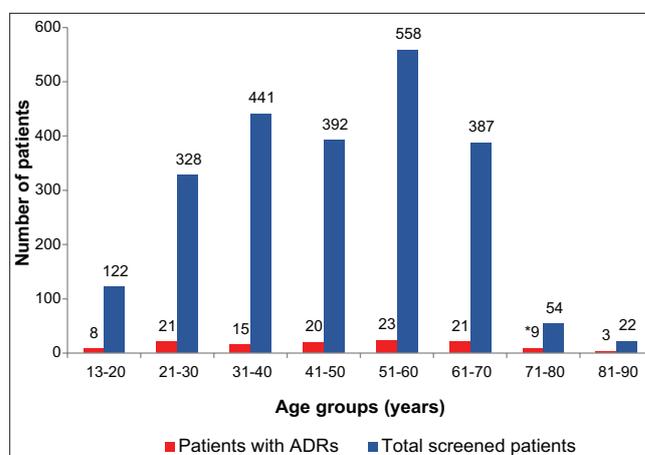


Fig. 1: Age-wise distribution of screened and enrolled patients who developed an adverse drug reaction at a tertiary care teaching hospital in Gujarat, India (n=120). Statistical significance was determined by Chi-square test. $P < 0.05$ was considered statistically significant. * $P < 0.05$ as compared to patients of all other age groups except 13–20 years and 81–90 years

Table 1: Drug groups suspected by experts and investigator for ADRs in enrolled patients (n=120)

Suspected drug group	Number of drugs suspected by experts (%)						
	Expert 1 (Total drugs – 209)	Expert 2 (Total drugs – 209)	Expert 3 (Total drugs – 155)	Expert 4 (Total drugs – 147)	Expert 5 (Total drugs – 259)	Expert 6 (Total drugs – 149)	Investigator (Total drugs – 190)
Antimicrobials	128 (61.2)	101 (48.3)	101 (65.2)	97 (66)	149 (57.5)	97 (65.1)	114 (60)
Drugs acting on renal system	18 (8.6)	16 (7.6)	16 (10.3)	14 (9.5)	21 (8.1)	16 (10.7)	17 (8.9)
Cardiovascular drugs	9 (4.3)	14 (6.7)	4 (2.6)	4 (2.7)	4 (1.5)	4 (2.7)	7 (3.6)
Drugs acting on respiratory system	8 (3.8)	5 (2.4)	2 (1.3)	2 (1.4)	6 (2.3)	2 (1.3)	3 (1.5)
Drugs acting on CNS	5 (2.4)	3 (1.4)	4 (2.6)	4 (2.7)	6 (2.3)	4 (2.7)	4 (2.1)
Drugs acting on GIT	4 (1.9)	12 (5.7)	0	0	24 (9.2)	0	5 (2.6)
Other miscellaneous drugs	37 (17.7)	58 (27.7)	28 (18.1)	26 (17.7)	49 (18.9)	26 (17.4)	43 (22.6)

ADR: Adverse drug reaction

Naranjo’s scale (70%), the French method (70%), and Karch and Lasagna Scale (65.8%). “Probable” (57.5%) and “Certain” (10%) associations were more frequent with Koh *et al.* scale (Table 3).

Agreement among different scales used by experts and investigator

A substantial agreement was observed among experts using the WHO-UMC scale (k=0.707), whereas none to the slight agreement was observed among experts using VAS (k=0.117) (Table 4).

A substantial agreement was present between Karch and Lasagna scale and the French method. A moderate agreement was present between the French method with Naranjo’s scale and WHO-UMC scale and also between Karch and Lasagna scale with Naranjo’s scale and WHO-UMC scale. Fair agreement was observed between VAS with the French method and Karch and Lasagna scale (Table 5).

Severity, preventability, and outcome of ADRs

A total of 101 (84.1%) ADRs were of moderate severity, 17 (14.1%) were mild, and two (1.6%) were severe. Ninety-two ADRs (77%) were not

preventable, 19 (15.8%) were probably preventable, and nine (7.5%) were definitely preventable. Ninety-one patients (75.8%) recovered completely from ADR during the hospital stay, 13 (10.8%) were recovering at the time of discharge, 2 (1.7%) died due to complications of ADRs, whereas the outcome of 14 (11.7%) ADRs was unknown.

Factors associated with disagreement among scales

Polypharmacy, serious ADRs, non-availability of laboratory data, and skin and subcutaneous tissue ADRs were found to be associated with disagreement among Karch and Lasagna scale, the French method, and WHO-UMC scale (Table 6).

DISCUSSION

CA is an integral part of pharmacovigilance. It is performed to identify important ADRs, to generate signals, and to evaluate the risk-benefit profile of drugs. Given its importance, an accurate CA is essential. While a number of tools are available for CA, few studies have evaluated their utility and agreement for ADRs occurring in indoor patients.

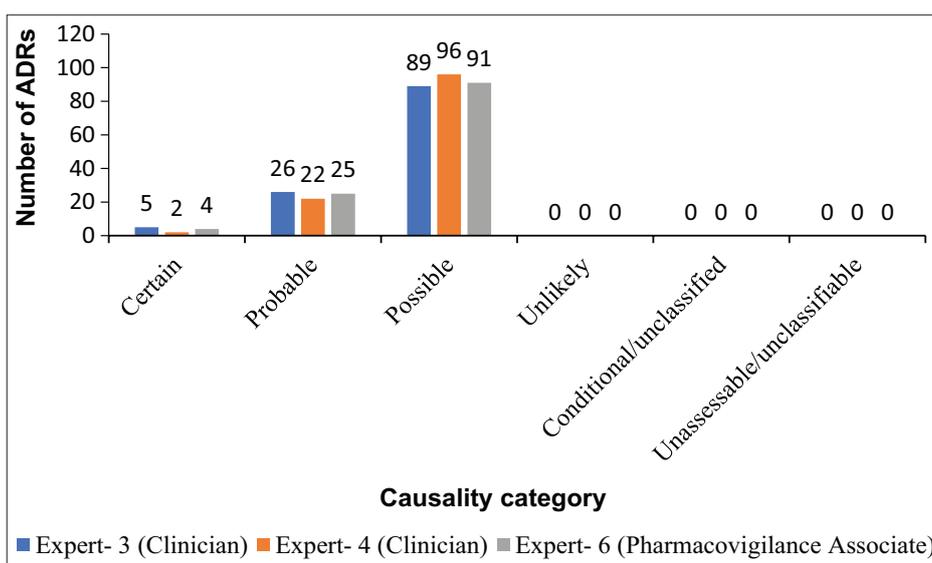


Fig. 2: Causality assessment for primary suspect drug and adverse drug reactions by experts using the WHO-UMC scale (n=120)

Table 2: Causality assessment by experts using Visual analogue scale

Causality category	Expert 1 (Clinician)		Expert 2 (Clinician)		Expert 5 (Pharmacologist)	
	Primary suspected drugs (%)	All suspected drugs (%)	Primary suspected drugs (%)	All suspected drugs (%)	Primary suspected drugs (%)	All suspected drugs (%)
Certain	3 (2.6)	3 (1.4)	0	0	11 (9.2)	11(4.2)
Likely	41 (35)	45 (21.5)	66 (63.5)	78(37.3)	45 (37.5)	45(17.4)
Plausible	35 (29.9)	52 (24.9)	15 (14.4)	31(14.8)	44 (36.7)	60(23.2)
Unassessable/Unclassifiable	17 (14.5)	30 (14.4)	9 (8.7)	24(11.5)	6 (5)	39 (15)
Doubtful	15 (12.8)	51 (24.4)	10 (9.6)	26(12.4)	9 (7.5)	55(21.2)
Unlikely	6 (5.1)	27 (12.9)	3 (2.9)	47(22.5)	5 (4.2)	47(18.1)
Excluded	0	1 (0.5)	1 (0.9)	3(1.4)	0	2(0.7)
Total	117 (100)	209 (100)	104 (100)	209 (100)	120 (100)	259 (100)

Experts 1 and 2 did not suspect any drug for causation of ADR in 3 and 16 cases, respectively.

Table 3: Causality assessment between the primary suspect drug and ADR by the investigator using four algorithms (n=120)

Causality category	Naranjo’s scale	Koh <i>et al.</i> Scale	French method	Karch and Lasagna scale
Certain/definite	2 (1.7%)	12 (10%)	Not applicable	5 (4.2%)
Probable	34 (28.3%)	69 (57.5%)	36 (30%)	36 (30%)
Possible	84 (70%)	39 (32.5%)	84 (70%)	79 (65.8%)
Doubtful/unlikely/conditional	0	0	0	0
Unrelated	Not applicable	Not applicable	Not applicable	0
Total	120 (100%)	120 (100%)	120 (100%)	120 (100%)

Table 4: Inter-rater agreement among experts using Visual analogue scale (VAS) and WHO-UMC scale (n=120)

Causality category	Inter-rater agreement in VAS			Inter-rater agreement in WHO-UMC scale		
	k	SE	Strength of agreement	k	SE	Strength of agreement
Certain – 1	0.180	0.545	None to slight	0.531	0.053	Moderate
Probable – 2	0.118	0.545	None to slight	0.716	0.053	Substantial
Plausible/ Possible – 3	-0.010	0.545	Worse	0.727	0.053	Substantial
Unassessable/unclassifiable, doubtful, unlikely and excluded – 4	0.210	0.545	Fair	NA	NA	NA
All categories	0.117	0.036	None to slight	0.707	0.047	Substantial

Kappa value (κ) was calculated using the Fleiss Kappa statistics, Number mentioned next to causality category was the code given to the respective causality category as described by Thaker *et al.* [6]

Table 5: Agreement of causality between different algorithms and scales used by investigator and experts (n=120)

Pair of algorithms	k	SE of kappa	95% CI	Strength of agreement
Naranjo's scale versus Koh <i>et al.</i> scale	0.232	0.060	0.116–0.349	Fair
Naranjo's scale versus Karch and Lasagna scale	0.577	0.077	0.427–0.728	Moderate
Naranjo's scale versus the French method	0.569	0.079	0.413–0.724	Moderate
Karch and Lasagna scale versus the French method	0.740	0.061	0.621–0.859	Substantial
Koh <i>et al.</i> scale versus the French method	0.181	0.056	0.072–0.290	None to slight
Koh <i>et al.</i> Scale versus Karch and Lasagna scale	0.275	0.060	0.146–0.404	Fair
VAS versus Naranjo's scale	0.112	0.048	0.017–0.206	None to slight
VAS versus Koh <i>et al.</i> scale	-0.301	0.05	(-0.128)–0.067	Worse
VAS versus the French method	0.309	0.063	0.186–0.433	Fair
VAS versus Karch and Lasagna scale	0.231	0.06	0.114–0.348	Fair
WHO-UMC versus Naranjo's scale	0.314	0.092	0.134–0.494	Fair
WHO-UMC versus Koh <i>et al.</i> scale	0.138	0.06	0.021–0.255	None to slight
WHO-UMC versus the French method	0.431	0.084	0.266–0.595	Moderate
WHO-UMC versus Karch and Lasagna scale	0.432	0.086	0.264–0.6	Moderate

Kappa value (κ) was calculated using Cohen's kappa statistics. VAS: Visual analog scale

Table 6: Association of various factors with disagreement among WHO-UMC scale, the French method and Karch, and Lasagna scale

Factor	Number of cases with the factor in Group A (n=80)	Number of cases with the factor in Group B (n=40)	P-value
Polypharmacy*	16 (20%)	26 (65%)	<0.0001
Severity of ADR	Mild – 12 (15%) moderate – 67 (83.7%) severe – one (1.25%)	Mild – five (12.5%), moderate – 34 (85%) severe – one (2.5%)	0.93 for mild 0.86 for moderate 0.61 for severe
Dechallenge	25 (31.3%)	12 (30%)	0.8
Serious ADRs*	5 (6.2%)	8 (20%)	0.048
Non-availability of laboratory data*	5 (25%)	5 (83%)	0.015
Similarity of ADR profile in suspected drugs	13 (16.2%)	10 (25%)	0.367
Affected SOC* Skin and subcutaneous tissue disorder	18 (22.5%)	18 (45%)	0.03
Similarity in suspected group of drugs	42 (52.5%)	21 (52.5%)	1

Polypharmacy: Concurrent use of more than five drugs. P-value was calculated using Chi-square test, *P<0.05 was considered statistically significant. ADR: Adverse drug reaction

In the present study, the incidence of ADRs was 5.2%. A lower incidence (2.12%) was reported by Doshi *et al.* [22] in a study of intensive monitoring of ADRs in indoor patients of two medical units of the same hospital. Rajpara and Kanani [23] also reported a low incidence (0.58%) of ADRs in a study conducted in indoor patients of a Tertiary Care Hospital in Vadodara, Gujarat. Differences in the prescribing pattern of drugs, as well as individual susceptibility, could have contributed to this discrepancy.

Patients >70 years of age were frequently affected by ADRs. However, the number of patients screened in these groups was substantially less compared to most groups. Elderly patients are more prone to ADRs due to factors such as multidrug therapy, changes in Pharmacokinetics and Pharmacodynamics of drugs [24]. However, Rajakannan *et al.* [25] reported a higher incidence of ADRs in the age group of 31–45 years (24.69%) compared to the age group of 61–75 years (20.19%) in a study conducted in South India. Further studies are recommended to evaluate the age-wise difference in the incidence of ADRs in different ethnic populations.

Infections (25.8%) were common in the study population. As a result, antimicrobials were frequently used and were the most common suspect

drugs during CA. Gastrointestinal, respiratory, and cardiovascular disorders were also frequently observed. The underlying disorder can produce signs and/or symptoms similar to ADR which can influence CA.

The common system-organs affected by ADRs were skin and subcutaneous tissue disorder (30%), gastrointestinal system disorder (24.1%), and metabolic and nutritional disturbances (15.8%). In the study by Doshi *et al.* [22], GI ADRs (27%) were most frequent and cutaneous ADRs (25%) were the most common cause of hospitalization. CNS ADRs (25.3%) were more frequent in the study by Rajpara and Kanani [23] followed by GI ADRs (14.9%) and skin and subcutaneous tissue disorders (13.8%). Variation in the pattern of ADRs can be due to differences in the pattern of drug use and individual susceptibility.

Following antimicrobial agents, drugs acting on the renal system were frequently suspected by experts and investigators since these drugs are often associated with metabolic and nutritional disorder ADRs. Other drug groups were less frequently suspected. Expert 5 demonstrated a tendency to suspect drugs acting on GI system frequently, whereas experts using the WHO-UMC scale did not suspect these drugs at all. This type of variation is expected in the expert judgment method as it

depends on the knowledge and experience of the assessor. Hence, inter-rater agreement is often poor.

A large variation was observed with regards to the total number of drugs suspected by experts using VAS. Increased communication among experts with a discussion of case safety reports can be employed to overcome such discrepancies. The variation was less among experts using the WHO-UMC scale since these experts suspected only one/two drugs for a given ADR. However, this tendency poses a risk of missing out on a rare drug event association.

A good agreement was observed with regards to the primary suspect drug in the majority of cases. However, experts using VAS showed disagreement in this regard in eight cases. Furthermore, experts 1 and 2 labeled some cases as 'not related to drug' as opposed to other experts, reflecting the inter-rater variation. A perfect agreement was observed with regards to the primary suspect drug among experts using the WHO-UMC scale, which can partly be attributed to suspicion of a less number of drugs.

Experts using VAS suggested "likely" and "plausible" associations between primary suspect drug and ADR most frequently. A variation, however, was observed with regards to two extremes of causality, that is, "certain" and "excluded". While clinicians (experts 1 and 2) chose "certain" relationship less frequently, the pharmacologist tended not to "exclude" a causal association with the suspect drug. Further studies are recommended to evaluate these tendencies in a larger number of experts. A fair number of "unassessable" and "doubtful" associations was also reported by experts. In the opinion of authors, the numerical nature of scale can be responsible for this finding as experts often judge the probability of a causal association in terms of percentage and not in terms of causality category. However, this opinion needs further evaluation. With regard to other suspect drugs, a fair number of "plausible" and "unassessable" associations were suggested by all experts, suggesting their suspicion regarding a possible causal role of these drugs. "Doubtful" and "unlikely" associations derived by experts suggested a tendency of not ruling out the drug causation possibility.

CA by experts using the WHO-UMC scale was more uniform as only three associations "Possible," "Probable," and "Certain" were derived upon. This can be attributed to the fact that only one/two most likely drugs were suspected by these experts. "Possible" association was most frequent, which suggested that ADRs presented with confounding factors, lack of dechallenge/negative dechallenge, and/or multiple suspect drugs. Sharma *et al.* [26] also reported a high proportion (73%) of "Possible" association in 200 ADRs by three experts using the WHO-UMC scale in a study conducted in Maharashtra, India. In a fair number of cases (22-29%), other drugs were also suspected to have a "Possible" causal association with the ADR.

Inter-rater agreement among experts using VAS was poor which suggests that this tool is not ideal to determine the causality of ADRs in indoor patients by a panel of experts. Arimone *et al.* [5] also reported none to the slight agreement ($k=0.2$) among five experts using VAS for CA of 150 drug-event pairs in 30 ADRs. On the other hand, a substantial agreement observed between experts using the WHO-UMC scale suggests that the criteria included in the scale are applied in a similar manner by experts even from different professional backgrounds and can be useful for CA in similar settings. Sharma *et al.* [26] also reported an almost perfect agreement ($k=0.89$) among three raters using the WHO-UMC Scale for CA of 200 ADRs in a study conducted in Maharashtra, India. Mouton *et al.* [27] also reported a substantial agreement ($k=0.61$) among four raters using the WHO-UMC scale for CA of 48 drug-event pairs.

CA by the investigator using algorithms revealed similar proportions of "Possible" and "Probable" associations with Naranjo's scale, the French method, and Karch and Lasagna scale. Lack of dechallenge/negative dechallenge, confounding factors and/or multiple suspect drugs

contributed to frequent "Possible" associations (65-70%) with these scales. Few "certain" associations were derived with Naranjo's scale and Karch and Lasagna scale and none with the French method since the latter avoids a "Definite" association [11]. "Probable" association was more frequently (57%) suggested with Koh *et al.* scale due to the difference in the scoring system as compared to other algorithms.

CA using the French method and Karch and Lasagna scale demonstrated a substantial agreement higher than any other pair of algorithms. Furthermore, both scales demonstrated moderate agreement with Naranjo's scale, which suggests the utility of these scales in the validation of CA with Naranjo's scale in similar settings. A moderate agreement of these two algorithms with the WHO-UMC scale was also observed; however, the agreement was lower than that with Naranjo's scale.

Polypharmacy, serious ADRs, non-availability of laboratory data and skin and subcutaneous tissue ADRs were found to be associated with disagreement among the WHO-UMC scale, the French method, and Karch and Lasagna scale. These factors need consideration during reporting and at the time of CA to ensure an accurate assessment.

Strengths of the study

A number of CA tools, that is, expert judgment combined with VAS and WHO-UMC scale and algorithms (Naranjo's scale, Koh *et al.* scale, the French method, and Karch and Lasagna Scale) were evaluated for agreement. Intensive monitoring of ADRs was performed in contrast to previous studies which have relied on spontaneously reported ADRs. Furthermore, these tools have not been evaluated for agreement with regards to ADRs occurring in indoor patients in India. Factors which can influence the agreement have also been evaluated.

CONCLUSION

WHO-UMC scale demonstrates a greater inter-rater agreement even among experts from different professional backgrounds, whereas VAS shows wide inter-rater variability, limiting its usefulness for CA in indoor patient settings. The French method and Karch and Lasagna Scale can be used for CA of ADRs in hospitalized patients as an alternative or adjunct to Naranjo's scale. Factors affecting agreement should be considered while reporting ADRs and during causality evaluation.

ACKNOWLEDGMENTS

We are thankful to Dr. A. N. Shah, former Professor and Head, Department of Medicine, Dr. Shivani A. Patel, Associate Professor, Department of Medicine, Dr. Kartikeya Parmar and Dr. Seema Sharma, Assistant Professors, Department of Medicine, Civil Hospital Ahmedabad, Dr. Jigar Kapadia, Assistant Professor, Department of Pharmacology, Mr. Alpesh Panseriya, former Pharmacovigilance Associate, B J Medical College, Ahmedabad, for their active participation in this research work as experts for CA.

AUTHORS' CONTRIBUTIONS

Dr. Chetan D. Parikh: Concept, literature search, data acquisition, data analysis, design, statistical analysis, manuscript preparation, manuscript editing, and manuscript review. Dr. Chetna K. Desai: Concepts, design, literature search, data analysis, statistical analysis, manuscript preparation, manuscript editing, and manuscript review. Dr. Jigar D. Kapadia: Literature search, data analysis, statistical analysis, manuscript preparation, manuscript editing, and manuscript review. Dr. Megha K. Shah: Concepts, design, literature search, data analysis, statistical analysis, manuscript preparation, manuscript editing, and manuscript review.

CONFLICTS OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

AUTHORS' FUNDING

The authors received no financial support for the research.

REFERENCES

- Kulkarni A, Bhatt A. Causality assessment: A casualty of compensation? *Perspect Clin Res* 2013;4:196-8.
- Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T. Systems analysis of adverse drug events. ADE prevention study group. *J Am Med Assoc* 1995;274:35-43.
- Arimone Y, Miremont-Salamé G, Haramburu F, Molimard M, Moore N, Fourrier-Réglat A, *et al.* Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions. *Br J Clin Pharmacol* 2007;64:482-8.
- World Health Organization, Uppsala Monitoring Centre. The Use of the WHO-UMC System for Standardized Case Causality Assessment. Geneva: World Health Organization; 2005. p. 2-7. Available from: <http://www.who-umc.org/Graphics/24734.pdf>.
- Arimone Y, Bégaud B, Miremont-Salamé G, Fourrier-Réglat A, Moore N, Molimard M, *et al.* Agreement of expert judgment in causality assessment of adverse drug reactions. *Eur J Clin Pharmacol* 2005;61:169-73.
- 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.
- 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.
- Koh Y, Li SC. A new algorithm to identify the causality of adverse drug reactions. *Drug Saf* 2005;28:1159-61.
- Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977;21:247-54.
- Bégaud B. Standardized assessment of adverse drug reactions: The method used in France. Special workshop--clinical. *Drug Inf J* 1984;18:275-81.
- Théophile H, Arimone Y, Miremont-Salamé G, Moore N, Fourrier-Réglat A, Haramburu F, *et al.* Comparison of three methods (consensual expert judgement, algorithmic and probabilistic approaches) of causality assessment of adverse drug reactions: An assessment using reports made to a French pharmacovigilance centre. *Drug Saf* 2010;33:1045-54.
- Koh Y, Yap CW, Li SC. A quantitative approach of using genetic algorithm in designing a probability scoring system of an adverse drug reaction assessment system. *Int J Med Inform* 2008;77:421-30.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;2:307-10.
- Liao JJ. Sample size calculation for an agreement study. *Pharm Stat* 2010;9:125-32.
- Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol* 2012;65:778-84.
- Hartwig SC, Siegel J, Schneider PJ. Preventability and severity assessment in reporting adverse drug reactions. *Am J Hosp Pharm* 1992;49:2229-32.
- Schumock GT, Thornton JP. Focusing on the preventability of adverse drug reactions. *Hosp Pharm* 1992;27:538.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378-82.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)* 2012;22:276-82.
- Doshi MS, Patel PP, Shah SP. Intensive monitoring of adverse drug reactions in hospitalized patients of two medical units at a tertiary care teaching hospital. *J Pharmacol Pharmacother* 2012;3:308-13.
- Rajpara AJ, Kanani NJ. An intensive monitoring of adverse drug reactions in indoor patients of medicine department at tertiary care teaching hospital: A single center, prospective, multisource observational study. *Natl J Physiol Pharm Pharmacol* 2019;9:1-8.
- Hilmer SN, McLachlan AJ, Le Couteur DG. Clinical pharmacology in the geriatric patient. *Fundam Clin Pharmacol* 2007;21:217-30.
- Rajakannan T, Mallayasamy S, Guddattu V, Kamath A, Vilakkthala R, Rao PG, *et al.* Cost of adverse drug reactions in a South Indian tertiary care teaching hospital. *J Clin Pharmacol* 2012;52:559-65.
- 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.
- 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.