

DEVELOPMENT AND VALIDATION OF A NOVEL METHOD FOR CAUSALITY ASSESSMENT USING SUSPECTED ADVERSE DRUG REACTIONS TO ANGIOTENSIN-CONVERTING ENZYME INHIBITORS

RAVI GOYAL¹, RUPINDER KAUR SODHI^{2*}, ANUJ GUPTA³

¹Department of Pharmacy, I. K. Gujral Punjab Technical University, Kapurthala, Punjab, India. ²Department of Pharmacology, Chandigarh College of Pharmacy, Mohali, Punjab, India. ³Community Health Centre, Nalagarh, Himachal Pradesh, India. Email: ccp.rupinder@cg.edu.in

Received: 06 June 2018, Revised and Accepted: 13 July 2018

ABSTRACT

Objectives: The study has been designed to develop, test reliability, and construct validity of a novel versatile causality assessment tool (VCAT) method.

Methods: The study included 427 literature case reports of angiotensin-converting enzyme inhibitors with suspected adverse drug reactions from 1990 to February 25, 2016. The causality of these cases was assessed independently by 3 raters in 3 phases of the study. Kappa (*k*) and intra-class correlation Intraclass Correlation Coefficient (ICC) were used to test reliability and validity of the VCAT method.

Results: Intra-rater reliability was high between Phase I and Phase III ($k=0.84-0.93$; % agreement: 92.3–96.3%). Inter-rater agreement was good in Phase I ($k=0.87-0.89$, % agreement: 93.7–94.1%, ICC: 0.975) and Phase III ($k=0.85-0.89$, % agreement: 93–94.4%, ICC: 0.973). Validity was proved by the high agreement observed between Phase I and Phase II ($k=0.78-0.94$; % agreement: 89.7–97.2%; $p<0.001$); and between Phase II and Phase III ($k=0.8-0.9$; % agreement: 90.2–94.8%; $p<0.001$).

Conclusion: VCAT method is a standardized causal assessment tool that gives valid and reproducible results. It has shown good agreement with the expert judgment method. This method may overcome the limitations enthrallled with existing methods of causality assessment.

Keywords: Angiotensin-converting enzyme inhibitors, Adverse drug reaction, Causality, Method, Rater.

© 2018 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.2018.v11i11.27749>

INTRODUCTION

Causality assessment is a pivotal tread of signal evaluation, to estimate the likelihood of drug in causing the adverse drug reaction (ADR) [1,2]. Monitoring of ADRs will aid in quantifying the risk associated with drugs. Non-drug causes acting as confounding factors may also be identified by careful assessment of the adverse event. Hence, causality assessment is crucial to decide the further course of the treatment in clinical trials and to validate signals in spontaneous reporting practice [2,3].

Numerous methods of causality assessment have been developed and reported over the years, which have been for broadly classified as: (1) Global introspection or expert judgment, (2) Bayesian approaches or probabilistic approaches, and (3) standardized assessment methods or algorithms [2,4]. Among these, global introspection is the most commonly used method where an expert's judgment is used to assign categories based on the association between the drug and the event. World Health Organization-Uppsala Monitoring Centre (WHO-UMC) assessment method is widely practiced. However, WHO-UMC method suffers from limitations of intra-rater and inter-rater variability [4-6]. Bayesian method uses prior/background information (epidemiological data) and reported event information to calculate posterior probability which represents the probability of the drug in causing the event [7]. These methods can simultaneously measure probability of multiple factors including drugs [6,7]. The major limitations of these methods include complexity, non-user friendliness, and time-consumption; hence, not very helpful in routine practice [4,8]. Among the standardized tools of causal assessment, Naranjo's algorithm is widely in practice. This method has a disadvantage of using the pathological data which might not always be reported/or available [4,9,10].

Considering the limitations encountered with the present evaluation methods, there is a need for a method which is comparatively

uncomplicated and encompasses standard parameters. This method may consider additional queries, incorporate the effects of drug interactions, class effect of drugs which are not quantified by methods in practice. This novel method should aim to attenuate inter-rater variability in addition to being reproducible and valid. The versatile causality assessment tool (VCAT) method developed during this study has the potential to be a versatile method that could provide reproducible results and minimize the variations among raters. Hence, through this study, we aim to establish the reliability and validity of the newly developed method of causality assessment using angiotensin-converting enzyme (ACE) inhibitors.

ACE inhibitors such as ramipril, enalapril, captopril, and fosinopril are a widely-used class of antihypertensive drugs. These are known to cause several adverse events; cough being the most common, often leading to drug discontinuation. Other serious events include angioedema and pancreatitis which could be life-threatening or fatal [11].

METHODS

VCAT description and instructions for use

VCAT may be a versatile tool to assess the causal association of the suspected adverse event to the drug. This method included five categories for causal assessment (certain, probable, possible, unclassified, and unlikely) and incorporates information on 10 parameters, reported in a case for suspectedness of the event. These parameters include: Time to onset (TTO), dechallenge (D), evidence in literature (L), class effect (C), dose dependency/response (DR), lab data/biopsy results supporting the event (LD), rechallenge (R), drug-drug interactions (DDI), medical history (H), and confounding drugs (CD). These parameters are explained in terms of the scope of the information as follows:

TTO: TTO is the temporal association between drug administration and event onset. Favorable response represents a time-related association;

unfavorable represents inconsistent timing for the drug to cause the suspected ADR; and unknown value represents the unassessable temporal association.

D: Improvement or recovery after discontinuation or dose reduction of the drug explains dechallenges. Favorable dechallenge denotes a positive outcome after drug discontinuation; vice versa in unfavorable; and unknown value denotes insufficient information. If the corrective treatment is given and the event resolves, dechallenge is taken as unknown as the role of the suspect drug cannot be accurately defined.

L: Evidence of the occurrence of similar events with the same drug reported in literature is important due to the scientific validity of the literature reports. Favorable denotes well-reported ADR with the drug in question; unfavorable denotes reports certainly specifying that the event in question is improbable with the suspect drug (this information will not generally be reported as no certainty can be provided); and unknown value denotes there is limited information in literature, or the drug is new.

C: Class effect is an important parameter to assess the role of the similar chemical structures in causing similar ADRs. Favorable effect represents well-reported ADR of a class of drugs; unfavorable effect represents certainty that the event in question is dubious with the class of drugs (this information will not generally be reported as no certainty can be provided); and unknown represents lack of information or newer drug with no drugs of the same class.

DR: Dose-response represents the effect of dose on the occurrence or intensity/severity of the adverse event. Some events occur particularly with the overdose of a drug while few conditions can aggravate with the increase in dose and can have a linear or exponential relationship. Favorable value symbolizes the role of dose in causing or aggravating the event; unfavorable symbolizes certainty that dose modifications will not cause/aggravate the event; and unknown value symbolizes lack of information about dose response to the ADR.

LD: Laboratory data supporting the event strengthens the evidence of the event in question. Favorable value signifies data supporting the event in question; unfavorable value signifies data opposing the event (this will impact the assessment adversely); and unknown value signifies lack of information or ambiguous information or reported without units.

R: Recurrence of the event on restarting the drug is a pivotal parameter to ascertain the role of the drug. Favorable value denotes positive rechallenge; unfavorable value signifies negative rechallenge; and unknown signifies lack of information. In general, rechallenge is not practiced due to ethical reasons, and the positive values are generally the results of accidental exposures/administrations or due to polypharmacy or self-medication.

DDI: DDIs can affect the causal assessment as the drug's plasma/tissue concentration can be modified by other drugs leading to adverse drug

events which can otherwise not be explained by the drug in question. Favorable value will have a negative impact on the likelihood of association and reduces the margin of association; unfavorable value signifies non-existence of DDI (this cannot be reported with certainty as extensive pharmacokinetic/pharmacodynamic studies are needed to ascertain); and unknown value signifies limited information about the concurrent use of the drugs in question or blood levels of both drugs or reported interactions.

H: Medical history is an etiological parameter that includes risk factors, concurrent conditions, family history, and underlying conditions. Favorable value will have a negative impact on the likelihood of association and reduces the margin of association; unfavorable value denotes that the patient is otherwise healthy with no medical conditions and this will increase the likelihood of association; and unknown value denotes limited information about the medical history of the patient.

CD: Concomitant drugs taken by the patient during the course of the drug in question, could act as confounding factors depending on their individual temporal association with the event. Favorable value signifies that the patient is taking other drugs that can act as alternative explanations and this will negatively impact the likelihood; unfavorable value signifies that the patient is not taking any other drugs and this will increase the likelihood of association; and unknown value signifies limited information about the use of concomitant medications.

Each of the 10 parameters can have favorable, unfavorable, or unknown value; and depending on this value, individual scoring is done. Favorable depicts positive values, unfavorable depict negative values and unknown depicts unassessable, limited or contradictory information. Each of the parameters is given a score depending on the information present in the case as presented in Table 1. In the end, all the scores are multiplied to reach a final score and categorized into 1 of the 5 categories.

Categorization based on Score range:

- Certain: >20
- Probable: >14–20
- Possible: >2–14
- Unclassified: 1–2
- Unlikely: <1.

The final score can range from a minimum of 0.02 to a maximum of 87.9. The minimum score might never be obtained since rechallenge will be not be performed if other causes explain the occurrence of the event; clearly indicating that other factors provide more likely explanations for the adverse event. The maximum value of 87.9 is also very rare; depicting high level of certainty that the event is related to the drug in question (none of the cases in this study achieved this value). All the parametric information used in assessment is based on the information at a specific point of time, i.e., in case of change of information (receipt of new information), parameter can change; hence, significantly affecting the case categorization.

Table 1: VCAT parametric table

Parameters/Condition	Favorable (Yes)	Unfavorable (No)	Unknown	Score
TTO	6	0.1	1	
D	1.5	1	1	
L	1.25	1	1	
C	1.25	1	1	
DR	1.25	1	1	
LD	1.25	0.5	1	
R	4	0.75	1	
DDI	0.5	1	1	
H	0.75	1	1	
CD	0.75	1	1	
Total score (by multiplication of all individual scores): $TTO \times D \times L \times C \times DR \times LD \times R \times DDI \times H \times CD$				

TTO: Time to onset; D: Dechallenge; L: Evidence in literature; C: Class effect; DR: Dose dependency/response; LD: Lab data/biopsy results supporting the event; R: Rechallenge; DDI: Drug-drug interactions; H: Medical history; CD: Confounding drugs. VCAT: Versatile causality assessment tool

In general, the score for each parameter ranges from 0.5 to 1.5. These values were adjusted after carefully considering extreme possible scenarios also keeping in mind the minimum and maximum values for the categories. By lowering the value, case category could change to a lower one, and by increasing the value, the category could be changed to an upper category. TTO is the foremost factor to decide the role of the drug; hence, more weightage has been given to it (6 for compatible and 0.1 for incompatible). Another important parameter, rechallenge has been given extra weightage in terms of score due to the significant impact of the parameters on the overall categorization of the assessment. During the preliminary analysis, this extra weightage given to these two parameters improved the results. Alternative etiologies (medical history and CD) if favorable, value is 0.75 for each parameter, reducing the overall score; and chances of the case of falling in a lower category are increased. If unfavorable, value is 1.5 for each parameter, increasing the overall score and chances of falling in a higher category are increased. For the cases with ADRs due to drug withdrawal, interpretation of the data is important as the withdrawal would relate to dechallenge and repeated withdrawal would relate to rechallenge. Other parameters remain the same.

Data collection

To identify all case reports related to "ADR with antihypertensive drugs," a literature search from the year 1990 to February 25, 2016 (search date) was performed in Embase and Medline databases (through Embase.com). Non-human/Preclinical studies were excluded. The drug class ("antihypertensive agent"/exp/mj) was restricted to major focus to retrieve the most relevant outputs in Embase.com.

A total of 7845 citations were retrieved and manually triaged from the indexing and abstract. All citations which were relevant for ADRs or were doubtful were shortlisted/retained. Only citations published in English language containing information related to the keywords of interest were retained, all foreign languages citations were excluded. The retained citations (approximately 5000) were checked for the availability of their full-text publications. After the availability of the full-text publications, first 1500 most recently, published articles were selected and were thoroughly reviewed for the occurrence of adverse events after the use of antihypertensive drugs. A total of 209 of these articles were identified to be irrelevant to our study due to the various reasons such as adverse events not occurring with antihypertensive drugs; lack of efficacy was reported; full-text articles were in foreign language and not English; no adverse event was reported; multiple unidentifiable patients; review articles; and duplicates. The remaining 1291 articles were screened to identify case reports with overdose (accidental and intentional) of antihypertensive drugs. A total of 160 articles reported overdose of antihypertensive drugs; which were removed from the analysis since the adverse drug event profile of a drug could differ substantially at an overdose. The remaining 1131 articles were included in the study and were identified with 1339 drug-event pairs as: (1) Few articles reported multiple patients and (2) few articles reported multiple events with different drugs in the same patient at different time intervals.

The selected 1339 drug-event pairs will be denoted as "cases" henceforth, which were further classified into the pharmacological classification of antihypertensive drugs: ACE inhibitors, beta-blocker, calcium channel blockers, diuretics, alpha-blocker, alpha-agonist, angiotensin receptor blockers, phosphodiesterase-5 inhibitors, and miscellaneous. Miscellaneous drugs included vasodilators, alkaloids, centrally acting antihypertensive, endothelin receptor antagonist, renin inhibitor, tyrosine inhibitor, and imidazoline antagonist.

Among these 1339 cases, 427 cases were identified with suspected adverse events occurring to ACE inhibitors. These cases were hence used for the reliability and validity testing of the newly developed VCAT method.

Procedure

The selected 427 cases were assessed by multiple raters at multiple time points to establish the validity and reliability of the VCAT method.

WHO-UMC method rates association between drug and event in one of the six categories (certain, probable, possible, unlikely, unclassified, and unclassifiable) and however, VCAT method has five categories (certain, probable, possible, unclassified, and unlikely). Literature case reports are clinically confirmed; hence, none of the cases were identified with insufficient information. Therefore, unclassified and unclassifiable categories (of the WHO-UMC method); and unclassified category (of the VCAT method), were not applicable to any of the cases. Thus, in both the methods, only four categories were applicable.

A total of 3 raters (1 physician with over 20 years of clinical experience, 1 pharmacologist with 15 years of experience, and 1 pharmacist with more than 7 years of experience in causality assessment) were involved in the assessment conducted in 3 phases.

Phase I: All 3 raters assessed 427 cases independently with VCAT method (referred to as 0-day reading) and were placed in 1 of the four applicable categories. Each of the 10 parameters could have favorable (yes), unfavorable (no) or unknown value, and depending on this value, scoring was done. In the end, all the scores were multiplied to reach a final score, and categorization was done based on the final score: Certain (>20), probable (>14–20), possible (>2–14), and unlikely (<1). Phase II: After 1 month, the 3 raters assessed the randomly reordered 427 cases with WHO-UMC method (referred to as 30-day reading) and were placed in 1 of the four applicable categories. This 1-month gap was given to allow the raters to forget the original judgments. Phase III: After another 2 months, the 3 raters again assessed all the randomly renumbered 427 cases with VCAT method (referred to as 90-day reading) and were placed in 1 of the four applicable categories. All the 3 raters met and discussed the results after each time interval to reach a consensus on the disagreement in the findings.

Reliability

Reliability testing was aimed to verify if different raters would reach at the same category of assessment for individual case reports at different time points. To test the reliability of the results with the VCAT method, assignments of the individual cases were statistically compared using Statistical Package for the Social Sciences (SPSS) 20.0. For intra-rater reliability, values of assignments with VCAT method at day 0 and day 90 were compared. For inter-rater reliability, assignments from all the 3-time intervals were compared (VCAT method at day 0, WHO-UMC method at day 30, and VCAT method at day 90). Intra-class coefficient (ICC) and Cronbach's Alpha (for internal consistency) were used for measurement of intra-rater and inter-rater reliability at different time intervals. Percentage agreement and kappa(k) were used to measure inter-rater reliability.

Validity

Validity testing aimed to check the proximity of the results to that produced by an external standard. The WHO-UMC method was used as the external standard for comparison with the VCAT method. The physician was referred in case of doubts on categorization. k statistics was used for validity testing to measure agreement between the methods and raters.

Materials required

All ADRs were categorized into appropriate system organ classes (SOCs) as per the Medical Dictionary for Regulatory Activities (MedDRA) (latest version). Drugs acting as confounding factors were identified using Micromedex, Martindale, Meyler's, and emc* (UK database for respective drug labels/Investigator Brochures). Risk factors/concurrent conditions/underlying disease acting as confounding factors were taken as per the discretion of the physician/existing literature. Inputs on laboratory data were taken as per the Common Terminology Criteria for Adverse Events. Information on class effect and past experience were referred from the existing literature and pharmacovigilance textbooks.

Statistics

All the data are presented using descriptive analysis. Statistical analysis was carried out using SPSS from IBM, statistical version 20.0. Reliability was calculated using the Cronbach's Alpha and ICC. *k* statistics was used for measurement of agreement between methods (ranging from -1 to +1). Per Landis and Koch [12], the values of *k* have been characterized as slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00).

The sampling technique used in the study was convenient sampling. This utility calculated the sample size required to estimate a proportion (prevalence=40% for hypertension) [13] with a specified level of confidence and precision. Based on the formula, $n = \frac{Z^2 X P (1-P)}{E^2}$ the

minimum required sample size was 369, where *Z* value was from standard normal distribution corresponding to the desired confidence level (*Z*=1.96 for 95% confidence interval), *p* (0.40) is expected true proportion, and *E* (0.05) is desired precision. We took a sample size of 427 to establish the validity and reliability of the VCAT method. All statistical tests were seen at the two-tailed level of significance ($p \leq 0.01$ and $p \leq 0.05$).

RESULTS

Of these 427 cases, 203 ADRs (47.5%) were in male patients, and 223 ADRs (52.2%) were in female patients, and the remaining 1 ADR (0.2%) was in unknown gender group. These 427 cases included 9 neonates (2.1%), 6 infant (1.4%), 11 children (2.6%), 7 adolescents (1.6%), 212 adults (49.7%), 181 elderly (42.4%), and age group was unknown in 1 case (0.2%). Death was reported in 16 (3.7%) of the 427 cases.

Event outcome was recovered in 243 cases (56.9%), recovering in 136 cases (31.9%), not recovered in 7 cases (1.6%), worsened in 1 case (0.2%), unknown in 37 cases (8.7%), and fatal in 3 cases (0.7%).

Table 2: Most frequently implicated drugs

Drug	n (%)
Lisinopril	127
Enalapril	118
Captopril	53
Ramipril	49
Perindopril	18

Table 3: Inter-rater agreement

Raters	Phase I				Phase II				Phase III			
	% agreement	k	Asymptotic standard error	p	% agreement	k	Asymptotic standard error	p	% agreement	k	Asymptotic standard error	p
Rater 1-Rater 2	94.1	0.883	0.023	0.001	96	0.921	0.019	0.001	93.4	0.866	0.025	0.001
Rater 2-Rater 3	93.7	0.869	0.025	0.001	91.8	0.827	0.028	0.001	93	0.856	0.026	0.001
Rater 1-Rater 3	94.4	0.888	0.022	0.001	93.4	0.867	0.024	0.001	94.4	0.887	0.023	0.001
ICC of reliability	0.975				0.973				0.973			

ICC: Intraclass correlation coefficient

Table 4: Intra-rater agreement

Raters	Phase I and II				Phase II and III				Phase I and III			
	% agreement	k	Asymptotic standard error	p	% agreement	k	Asymptotic standard error	p	% agreement	k	Asymptotic standard error	p
Rater 1	97.2	0.945	0.016	0.001	94.8	0.898	0.021	0.001	96.3	0.926	0.018	0.001
Rater 2	91.1	0.816	0.029	0.001	92.3	0.840	0.027	0.001	92.3	0.840	0.027	0.001
Rater 3	89.7	0.782	0.031	0.001	90.2	0.795	0.030	0.001	93	0.856	0.026	0.001

SOC wise distribution of these 427 cases, per MedDRA 20.1, is presented in Fig. 1. Maximum adverse events were related to the following SOCs: Skin and subcutaneous tissue disorders: 205 cases (48%); Gastrointestinal disorders: 55 cases (12.9%); and Metabolism and nutrition disorders: 29 cases (6.8%).

Angioedema was the most frequently reported (150 [35.1%]) event with ACE inhibitors. The most commonly implicated drugs are presented in Table 2.

Reliability

In Phase I (VCAT at 0-day); inter-rater percentage agreement varied from 93.7% to 94.1% with *k* ranging from 0.87 to 0.89 and ICC of reliability at 0.975. In Phase II (WHO-UMC at 30-day); inter-rater percentage agreement ranged from 91.8% to 96% with *k* varying from 0.83 to 0.92 and ICC of reliability at 0.973. In Phase III (VCAT at 90-day); inter-rater percentage agreement varied from 93% to 94.4% with *k* ranging from 0.85 to 0.89 and ICC of reliability at 0.973 ($p < 0.001$). ICC of reliability suggests a high level of reproducibility. In Phase I and Phase III, a highest agreement was observed between Rater 1 and Rater 3 while in Phase II, a highest agreement was observed between Rater 1 and Rater 2 (Table 3).

Intra-rater reliability after comparing the results through VCAT method in Phase I (VCAT at 0-day) and Phase III (VCAT at 90-day) was high with a percentage agreement ranging from 92.3% to 96.3% with *k* ranging from 0.84 to 0.93 (Table 4). Rater 1 had shown maximum agreement followed by Rater 3 and Rater 2, respectively.

Validity

Agreement between Phase I (VCAT at 0-day) and Phase II (WHO-UMC at 30-days) varied from 89.7% to 97.2% with *k* ranging from 0.78 to 0.94. Agreement between Phase II and Phase III (VCAT at 90-day) ranged from 90.2% to 94.8% with *k* varying from 0.8 to 0.9 ($p < 0.001$). Rater 1 had shown the maximum agreement in all the comparisons (Table 4).

DISCUSSION

Data indicates that VCAT method has shown a high reproducibility between different raters and at different intervals of time. It also shows good agreement with a WHO-UMC method which is a widely accepted method of causal association. ICC of reliability as a measure of inter-rater reproducibility was observed at 0.97 in Phase I and Phase III where results were obtained using VCAT method.

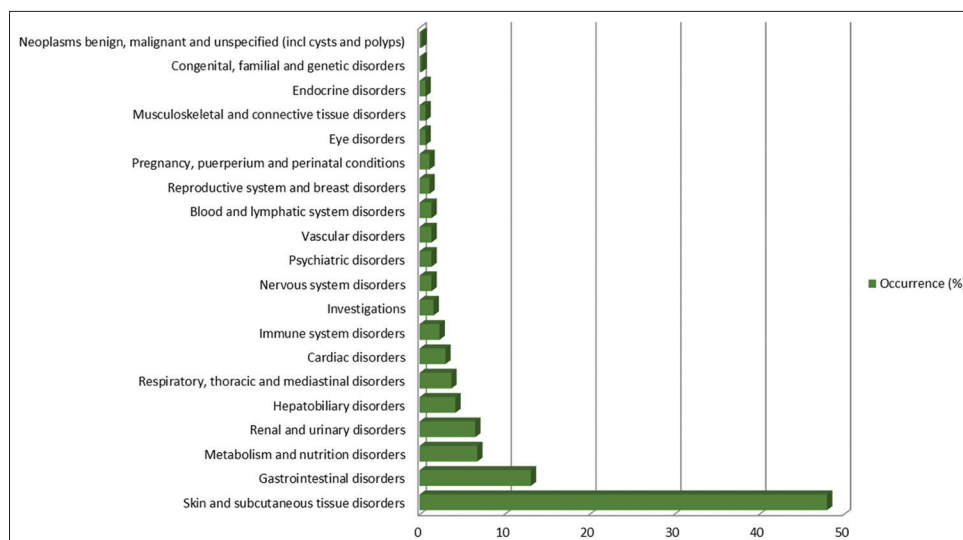


Fig. 1: SOC wise classification of ADRs

Reproducibility of the instrument is pivotal as it allows the results of different raters to be compared at different times and places. High agreement ($k=0.84$ to 0.93 ; % agreement: 92.3% – 96.3%) was observed with VCAT method (Table 4) when comparisons were drawn between two assessments by the same method as day 0 (in Phase I) and day 90 (in Phase III), indicating intra-rater reliability of the VCAT method. High inter-rater agreement was observed in Phase I ($r=0.975$; $k=0.87$ – 0.89 ; % agreement: 93.7% to 94.1%) and Phase III ($r=0.973$, $k=0.85$ – 0.89 ; % agreement: 93% – 94.4%) during comparison among all the raters (Table 3); indicating inter-rater reproducibility of the VCAT method.

Measurement of validity against a standard scale is also very important to establish the acceptability of the values from any new scale or method. We had used WHO-UMC method as a standard tool to establish the validity of our method. Content validity of the VCAT method was indicated with low scoring in cases with certain presence of alternative etiologies and high scoring in cases with true ADR. The concurrent validity of our method was established by the fact that very good agreement was seen between VCAT method and the WHO-UMC method in both the comparisons (Phase I vs. Phase II and Phase II vs. Phase III). All the raters in both the comparisons had a good agreement with Rater 1 showing the maximum agreement in both. This could be possibly because Rater 1 was a clinician with good clinical experience. Moreover, Rater 2 (pharmacologist) and Rater 3 (pharmacist) had shown good agreement between both the methods; indicating that VCAT method can be used effectively by non-clinicians as well. This signifies extension of the utility of this method to all health-care professionals. In a study by Naranjo *et al.* [9], non-experts have shown improvement in reliability with the use of the standardized tool.

High reproducibility was also observed in Phase II (Table 3), where WHO-UMC method was used (percentage agreement ranged from 91.8% to 96% with k varying from 0.83 to 0.92 and ICC of reliability at 0.973). Although the categories are well defined, yet some raters may differ on certain parameters leading to a difference in assessments. This finding is interesting in the view of a study by Naranjo *et al.* [9], where ICC of reliability improved from 0.49 (value obtained by an individual expert judgment without using any definite scale/method) to 0.92 when ADR probability score method was used.

A high agreement between raters could be expected by chance, but this seemed doubtful as the data were analyzed in multiple ways, and the categorization of the cases was not narrow. Another likely explanation could be the selection of straightforward/simple case scenarios; this

also is improbable as literature case reports have explicit data and is generally cumbersome to assess. Another hypothesis which seems plausible is the role of the assessment tool in giving reproducible results even when used by non-clinicians.

VCAT method improves the agreements, and at the same time, the disagreements in parameters among raters can also be identified. These disagreements can be important in specific scenarios to answer difficult questions in the case of new drugs and new symptoms. In our study, the majority of the disagreements among raters was due to differences in the alternative etiologies, i.e., identifying the confounding medical history and concomitant drugs taken by the patient; and dechallenge in context of corrective treatment given for the event. Pere *et al.* [14], in a study, had indicated that main causes of disagreement were a temporal association, alternative explanations, and dechallenge. Studies by Naranjo *et al.*, 1981 [9], and Hutchinson *et al.* [15] reported that most of the disagreements among raters were due to the alternative explanations as the non-clinicians could miss out on some of the confounding factors. Hypertension and other comorbid disease including diabetes predispose patients to increased risk of ADRs due to the sub-optimal functioning of the organ systems. Other factors could be patient age and multiple drug therapy. These factors acted as confounding factors in a significant number of cases in the possible category in our study.

Studies have suggested that quantification of the parameters cannot be done in the WHO-UMC method; hence, there is no empirical rationale for the categorization of an ADR. However, with the algorithms/scales, classification of the categories is based on the empirical values that could quantify the probability of the ADR causes. Interestingly, high level of agreement between VCAT and WHO-UMC method signifies results comparable to that of global introspection method with the added benefit of giving reproducible results even when applied by non-clinicians. This agreement could also be attributed to some extent to the retrospective nature of the literature reports used in the study which is medically validated, and the assessor is unable to discuss the findings with the prescriber. Similar kind of findings was obtained in a study by Kane-Gill *et al.* [10], where the high agreement was observed in the retrospective phase ($k=0.794$) compared to prospective phase ($k=0.635$).

Inter-rater and intra-rater variations are the major limitations in all the methods of causality assessment. A method that could overcome this limitation and gives reproducible results is needed. Use of conventional definitions could give varying results as indicated by studies by Naranjo *et al.* [9], Karch *et al.* [16], Blanc *et al.* [17], and Koch-Weser *et al.* [18]. Causality assessment uses the available data in a systematic manner and helps to enhance the scientific evaluation. Inevitably, there could never be

a perfect agreement between different raters and different methods due to variation in subjective thinking, but it can be minimized so that a basic level of comparison could be drawn. The important characteristic of a method is the identification of the underlying cause of the event that might be drug or non-drug. This understanding will help to reduce the burden of the ADRs from the health-care system by evaluation of safety signals which will eventually help identify and reduce the risks to public health [19].

Clinical judgment is also essential while using the standardized tools as some of the questions can be unanswered by a scale that might require medical judgment. Training is needed before using the assessment tools, and care should be taken to minimize or avoid the error, to get good quality assessments. The VCAT method can be used for a wide variety of reports from both solicited (clinical trials, patient access programs, and patient-oriented programs, etc.) and unsolicited (spontaneous, literature, and patient registries, etc.) sources. In the case of multiple CD, VCAT should be applied to find the drug with the highest score which will have more probability in causing the adverse event. However, it is expected that our method may not solve some of the complex or extreme cases where a clinical experience might be required for certain adjustments.

CONCLUSION

VCAT is a standardized tool of causal assessment that gives reproducible results when compared with the WHO-UMC method and improves the quality of the assessment. We suggest using VCAT method for causality assessment in both prospective and retrospective studies to reach a valid and reproducible result. Results from the VCAT method may also be compared with other causality assessment methods to further confirm its validity. The applications of this method may also be implied to other pathological conditions and drug classes. Its use in clinical practice will pave the way for further improvements.

ACKNOWLEDGMENTS

We are thankful to Mr. Rajesh Kumar, Statistical officer, Kalpana Chawla Medical College, Karnal, Haryana, India, for carrying out the statistical analysis and providing his valuable inputs. We are grateful to IK Gujral Punjab Technical University, Kapurthala, for the needed support.

AUTHOR'S CONTRIBUTIONS

The work was carried out by Ravi Goyal, under the guidance of Dr. Rupinder Kaur Sodhi and Dr. Anuj Gupta. Both, Dr. Rupinder Kaur Sodhi, and Dr. Anuj Gupta, have also contributed to the causality analysis at different time intervals.

DECLARATIONS OF INTEREST

The authors do not have any financial conflicts with any organization.

REFERENCES

1. Arimone Y, Miremont-Salame G, Haramburu F, Molimard M, Moore N, Fourier-Reglat A, *et al.* Inter-expert agreement of seven criteria in

- causality assessment of adverse drug reactions. *Br J Clin Pharmacol* 2007;64:482-8.
2. Belhekar NM, 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.
3. Mathew E, Chandrika C, Karanath PM, Srinivasa R. A prospective observational study on prescribing trends and adverse drug reactions in stroke patients. *Int J Pharm Pharm Sci* 2017;9:25-30.
4. Agbabiaka TB, Savovic J, Ernst E. Methods for causality assessment of adverse drug reactions: A systematic review. *Drug Saf* 2008;31:21-37.
5. WHO-Uppsala Monitoring Centre. The use of the WHO-UMC System for Standardised Case Causality Assessment. Available from: http://www.who.int/medicines/areas/quality_safety/safety_efficacy/WHOcausality_assessment.pdf. [Last cited on 2018 Mar 20].
6. Hire RC, Kinage PJ, Gaikwad NM. Causality assessment in pharmacovigilance: A step towards quality care. *Sch J App Med Sci* 2013;1:386-92.
7. Hutchinson TA, Dawid AP, Spiegelhalter DJ, Cowell RG, Roden S. Computerized aids for probabilistic assessment of drug safety I: A spreadsheet program. *Ther Innov Regul Sci* 1991;25:29-39.
8. Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. *Drug Saf* 1997;17:374-89.
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. Kane-Gill SL, Forsberg EA, Verrico MM, Handler SM. Comparison of three pharmacovigilance algorithms in the ICU Setting: A retrospective and prospective evaluation of ADRs. *Drug Saf* 2012;35:645-53.
11. Parish RC, Miller LJ. Adverse effects of angiotensin converting enzyme (ACE) inhibitors. An update. *Drug Saf* 1992;7:14-31.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
13. WHO: Global Health Observatory (GHO) Data. Raised Blood Pressure. Available from: http://www.who.int/gho/ncd/risk_factors/blood_pressure_prevalence_text/en/. [Last cited on 2018 Feb 06].
14. Pere JC, Begaud B, Haramburu F, Albin H. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986;40:451-61.
15. Hutchinson TA, Leventhal JM, Kramer MS, Karch FE, Lipman AG, Feinstein AR. An algorithm for the operational assessment of adverse drug reactions. II. Demonstration of reproducibility and validity. *JAMA* 1979;242:633-8.
16. Karch FE, Smith CL, Kerzner B, Mazzullo JM, Weintraub M, Lasagna L. Adverse drug reactions: A matter of opinion. *Clin Pharmacol Ther* 1976;19 5 Pt 1:489-92.
17. Blanc S, Leuenberger P, Berger JP, Brooke EM, Schelling JL. Judgments of trained observers on adverse drug reactions. *Clin Pharmacol Ther* 1979;25 5 Pt 1:493-8.
18. Koch-Weser J, Sellers EM, Zcest R. The ambiguity of adverse drug reactions. *Eur J Clin Pharmacol* 1977;11:75-8.
19. Gauravkumar S, Mukeshkumar P, Bhanwarlal J. Safety signal detection of cardiac disorders adverse drug events for azithromycin in pediatric population using health Canada adverse event reporting system database. *Int J Pharm Pharm Sci* 2017;9:80-4.