Regulatory Toxicology and Pharmacology 59 (2011) 1–7

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Commentary

Regulatory causality evaluation methods applied in kava hepatotoxicity: Are they appropriate?

Rolf Teschke *, Albrecht Wolff

Department of Internal Medicine II, Division of Gastroenterology and Hepatology, Klinikum Hanau, Teaching Hospital of the Johann Wolfgang Goethe-University of Frankfurt/Main, Germany

ARTICLE INFO

Article history:
Received 20 July 2010
Available online 18 September 2010

Keywords:
Kava
Kava hepatotoxicity
Herbal hepatotoxicity
Drug-induced liver injury
Herbs induced liver injury
Hepatotoxicity
Toxic liver disease
Toxic liver injury
Causality assessment

ABSTRACT

Since 1998 liver injury has been assumed in some patients after the use of kava (*Piper methysticum* G. Forster) as an anxyolytic herbal extract, but the regulatory causality evaluation of these cases was a matter of international and scientific debate. This review critically analyzes the regulatory issues of causality assessments of patients with primarily suspected kava hepatotoxicity and suggests recommendations for minimizing regulatory risks when assessing causality in these and other related cases. The various regulatory causality approaches were based on liver unspecific assessments such as ad hoc evaluations, the WHO scale using the definitions of the WHO Collaborating Centre for International Drug Monitoring, and the Naranjo scale. Due to their liver unspecificity, however, these causality approaches are not suitable for assessing cases of primarily assumed liver related adverse reactions by drugs and herbs including kava. Major problems emerged trough the combination of regulatory inappropriate causality assessment methods with the poor data quality as presented by the regulatory agency when reassessment was done and the resulting data were heavily criticized worldwide within the scientific community. Conversely, causality of cases with primarily assumed kava hepatotoxicity is best assessed by structured, quantitative and liver specific causality algorithms such as the scale of the CIOMS (Council for International Organizations of Medical Sciences) or the main-test as its update. Future strategies should therefore focus on the implementation of structured, quantitative and liver specific causality assessment methods as regulatory standards to improve regulatory causality assessments for liver injury by drugs and herbs including kava.

1. Introduction

Herbal hepatotoxicity is a rare but potentially life-threatening disease and requires special attention for both treating the affected patients and ascertaining a sound diagnosis (Seeff, 2007; Navarro, 2009). Of particular importance is the early suspicion and collection of all relevant data of the case under consideration to facilitate subsequent causality assessment (Teschke and Bahre, 2009). As difficult as it may be to unequivocally establish drug-induced liver injury of conventional synthetic drugs, it is even more difficult to implicate herbal products for the many reasons such as product purity, product contamination and adulteration (Borrelli and Ernst, 2008; Health Canada, 2010). In addition, causality evaluation may be confounded by various inconsistencies and factors such as lack of a temporal association; missing definitions of the adverse reaction; inappropriate treatment modalities with high product doses and prolonged use; missing challenge and dechallenge data; alcohol consumption; alternative diagnoses; comorbidity; and coadministration with other synthetic drugs, herbal drugs and dietary supplements containing a variety of other herbs as mixture (Teschke et al., 2009a,b). Other challenging issues commonly recognized are poor qualities of data primarily collected by the treating physicians (Teschke et al., 2009b) and inadequate regulatory data presentation (Liss and Lewis, 2009). Taking these limitations into account, various open questions remain as to whether the use of an herb was really causally related to any liver disease.

Herbal hepatotoxicity by the use of the anxiolytic herb kava (pepper family Piperaceae, *Piper methysticum* G. Forster) is a particular challenging issue (Schmidt et al., 2005; WHO, 2007). Thorough analyses are available as reviews regarding its clinical aspects (Teschke, 2010a) and pathogenetic factors (Teschke, 2010b). The present review will focus on the regulatory shortcomings of data presentation and causality evaluation which are of common interest with respect to pharmacovigilance considerations.

* Corresponding author. Address: Department of Internal Medicine II, Klinikum Hanau, Teaching Hospital of the Johann Wolfgang Goethe-University of Frankfurt/Main, Leimenstrasse 20, D-63450 Hanau, Germany. Fax: +49 6181 2964211.
E-mail address: rolf.teschke@gmx.de (R. Teschke).

0273-2300/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2010.09.006

2. Regulatory data presentation

Expectations are high when a regulatory agency issues a withdrawal of an herbal drug such as kava from the market and presents the pharmacovigilance data of the cases (BfArM, 2002), especially when problems of toxic liver disease presumably associated with kava extracts have to be discussed regarding causal relationship, role of solvents for aqueous, ethanolic and acetonic extracts, kava raw material, comedication, dosage and duration of intake, impurities, and adulteration (WHO, 2007; Teschke, 2010a,b). There was worldwide interest and analysis, and the general conclusion was reached that the data quality of the regulatory presented cases with primarily suspected kava hepatotoxicity was poor and inappropriate (Denham et al., 2002; Schulze et al., 2003). Despite international criticisms and the requests of various scientific groups to provide additional data (Denham et al., 2002; Teschke et al., 2003; Schmidt et al., 2005), the regulatory agency failed to follow these suggestions (BfArM, 2005). No major regulatory attempts have been made to present, for instance, results concerning exclusion of non kava and non drug causes (Teschke et al., 2003), although these and other data have basically been available and were published later on with thorough analyses in scientific journals (Teschke et al., 2008a; Teschke and Wolff, 2009; Teschke, 2010a). The regulatory information of the patients was also selective, incomplete and thereby inadequate (Teschke and Wolff, 2009). It therefore appears that the regulatory data presentation in general was disappointing for the scientific community (Denham et al., 2002; Schulze et al., 2003; Teschke et al., 2003; Schmidt et al., 2005; Teschke and Wolff, 2009).

Spontaneous signaling programs carried out by regulatory agencies are usually based on accumulated reports that meet a case definition, sometimes referred as signal generation. In recent years much work is being done on the use of data mining methods for signaling, procedures that are independent of content, solely based on statistical disproportionality. The use of these regulatory causality methods may be helpful in the field of herbal pharmacovigilance, but evidence was not presented that these methods had actually been applied for regulatory assessment in cases of suspected kava hepatotoxicity (BfArM, 2002). Prior to pharmacovigilance assessment, however, an exhaustive evaluation of each individual case is required, since quality of causality assessment is more important than quantity of poorly assessed cases (Teschke et al., 2009c).

3. Ad hoc causality assessment

In 2002, the regulatory ban of kava was based not only on poor data but also on a narrative causality assessment, suggesting obviously some kind of an ad hoc causality approach by guilt by association (BfArM, 2002). There is no question that the use of an ad hoc causality assessment method for cases with liver injury is highly debatable, since this approach is inaccurate and lacks liver specificity (Kaplowitz, 2001; Teschke and Wolff, 2009).

Various items are usually considered essential for this type of assessment but certainly open for discussion (Table 1); in particular, there is no universally accepted description given for this method or its usage. Having ruled out nondrug causes, a distinction of a probable, possible, and unlikely causality is often used (Kaplowitz, 2001). A probable causality is usually assigned when the manifestation of liver disease, temporal association, and dechallenge response fit the typical signature of the drug in question. A possible causality is assigned when one of these parameters is not typical, the drug is not known to cause the reaction, or so rarely that it is difficult to distinguish from background, or an alternative cause is less or equally plausible. An unlikely causality is assigned when most of the features are atypical or an alternative cause is more plausible. Obviously, this simple distinction between levels of probability of assigning causality cannot be accurately and reproducibly applied to every case and is likely to foster disagreement among experts. In practice, this ad hoc approach is attempting to give a "yes, no, or may be" answer to a diagnosis without a gold standard. It has been pointed out that in the absence of liver specific causality assessment methods there has been no sound basis for determining the likelihood that an episode of hepatitis represents a drug-related reaction (Lee, 2003; Gunawan and Kaplowitz, 2004; Maddrey, 2005. The inaccuracy of the ad hoc causality approach is highlighted by a high rate of diagnoses missed upon assessment, and the correct diagnoses became evident upon subsequent thorough analysis including also quantitative assessment methods (Aithal et al., 1999). Missed diagnoses were not restricted to primarily suspected drug-induced liver injury (Aithal et al., 1999; Andrade et al., 2006; García-Cortés et al., 2008; Teschke et al., 2008b) but included also herbal hepatotoxicity (Teschke et al., 2008a, 2009a,b). Under these conditions, a patient with an incorrectly diagnosed disease is inappropriately being treated, whereas the real existing disease lacked a specific treatment in time; this delay may result in a deleterious outcome.

Table 1
Ad hoc causality assessment.

Items
1. Signature of clinical manifestation
2. Latency period
3. Dechallenge
4. Definitive exclusion of alternative causes
5. Risk factors
6. Alcohol
7. Other diseases
8. Track record of the drug

Details are derived from Kaplowitz (2001), Gunawan and Kaplowitz (2004), and Maddrey (2005).

Not presenting any criteria used for the assessing method is quite unusual for a regulatory agency (BfArM, 2002) and was unexpected but possibly explained by the poor data quality (Teschke and Wolff, 2009). Under the latter conditions, the initial causality assessments of scientific groups have also been achieved only on an ad hoc basis (Denham et al., 2002; Teschke et al., 2003; Schmidt et al., 2005), in accordance with other regulatory agencies such as the MCA (Medicines Control Agency) or EMA (European Medicines Agency, formerly EMEA) (Teschke et al., 2003; Schmidt et al., 2005; Teschke and Wolff, 2009). In all patients with primarily suspected kava hepatotoxicity, the regulatory assessment yielded various levels of causality categories for kava: causality was highly probable, probable, probable/possible, and possible in 2, 14, 2, and 7 patients, respectively (BfArM, 2002; Teschke et al., 2008a), despite shortcomings regarding regulatory data presentation, selection and major deletions (Teschke and Wolff, 2009). Based on identical regulatory presented case data and identical ad hoc causality assessments, the high regulatory causality ranking for kava was not reproducible; rather than low graded causality was suggestive, and this in only a few patients, as evaluated by MCA, EMEA, and various scientific groups (Denham et al., 2002; Schulze et al., 2003; Teschke et al., 2003; Schmidt et al., 2005; Teschke and Wolff, 2009). As expected, the combination of poor data quality with inappropriate causality assessment methods led to unacceptable results. In accordance with this impression is the high rate of diagnoses missed by the regulatory ad hoc assessment of patients with primarily assumed kava hepatotoxicity (Teschke et al., 2008a; Teschke, 2010a). It is clear that missed diagnoses are in no way acceptable, neither for the section of pharmacovigilance nor for physicians treating patients with primarily assumed liver injury

by drugs and herbs; missed diagnoses may easily convert to legal problems (Teschke et al., 2008b). It therefore appears that the regulatory approach of an ad hoc causality assessment was not convincing and should have been avoided in view of the shortcomings and the expected criticisms.

4. WHO scale

In 2005, the German regulatory agency informed the scientific community about the advantages of the liver unspecific WHO method applied for assessment of its regulatory cases of suspected kava hepatotoxicity (BfArM, 2005). The definitions of the WHO Collaborating Centre for International Drug Monitoring are used by this method (WHO, 2000), being the WHO scale in short (Table 2). Regulatory argumentation included the opinion that the WHO scale is an established tool for causality assessment for general adverse drug reactions (BfArM, 2005), but this does certainly not apply to hepatotoxic reactions. At least in scientific reviews and books dealing in depth with causality assessment methods to evaluate cases of toxic liver injury, the WHO scale is commonly not mentioned and therefore out of discussion (Zimmerman, 1999; Andrade et al., 2004; Teschke et al., 2008b), not supporting views to the contrary and the regulatory preference to use this particular scale for the kava cases (BfArM, 2005). Thus, in the context of liver injury the WHO scale is outside the expert's considerations.

Analyzing the details and going through the WHO scale point by point (Table 2), it is evident that the items of this scale are not only liver unspecific but also extremely vague. There is, for instance, also lack of any time frame for the challenge and dechallenge period, and criteria how to exclude other causes are missing. Consequently, the regulatory used WHO scale was obviously not in a position to recognize missed diagnoses in cases of primarily assumed kava hepatotoxicity (BfArM, 2005), in contrast to other successful approaches that easily discovered faulty diagnoses (Teschke et al., 2008a; Teschke, 2010a). This scale is therefore not suitable for assessing cases of suspected liver injury by drugs and herbs including kava. Certainly, the WHO scale may have its place to assess causality of diseases of organs other than the liver, but it is obsolete for liver injury.

Table 2
WHO scale.

Items
1. Certain causality • Event or laboratory test abnormality, with plausible time relationship to drug intake • Cannot be explained by disease or other drugs • Response to withdrawal plausible (pharmacologically, pathologically) • Event definitive pharmacologically or phenomenologically (i.e. an objective and specific medical disorder or a recognized pharmacological phenomenon) • Rechallenge satisfactory, if necessary
2. Probable causality • Event or laboratory test abnormality, with reasonable time relationship to drug intake • Unlikely to be attributed to disease or other drugs • Response to withdrawal clinically reasonable • Rechallenge not required
3. Possible causality • Event or laboratory test abnormality, with a reasonable time relationship to drug intake • Could also be explained by disease or other drugs • Information on drug withdrawal may be lacking or unclear
4. Unlikely causality • Event or laboratory test abnormality, with a time relationship to drug intake that makes a relationship improbable (but not impossible) • Disease or other drugs provide plausible explanations
5. Unclassified causality • Event or laboratory test abnormality • More data for a proper assessment needed, or • Additional data under examination
6. Unassessable causality • Report suggesting an adverse reaction • Cannot be judged because information is insufficient or contradictory • Data cannot be supplemented or verified

Details are derived from WHO (2000).

It is interesting to note that regulatory discussions and appraisals of the WHO scale in association with the regulatory kava cases finally led to possible/probable causalities for kava attributed to two patients (BfArM, 2005), but the WHO scale does not allow for this kind of intermediate causality category (Table 2) (WHO, 2000). With respect to the regulatory assessed intermediate causality categories, it was argued that there may be sometimes no clear cut separation between two different category levels (BfArM, 2005) which strengths the argument that the use of this scale lacks accuracy despite some defined criteria (Table 2). When at least one item is missing required for a particular level of probability, the level below is then the correct category not requiring intermediate levels. Thus, the inconsistency of reporting intermediate causality categories remained to be solved.

Additional inconsistencies emerged when in the kava WHO report identical regulatory kava cases were reassessed using the WHO scale (WHO, 2007), identical to that one used by the regulatory agency (BfArM, 2005). When the results of causality assessments of kava cases were compared, there were major inconsistencies of the causality ratings between the regulatory agency (BfArM, 2005) and the WHO report (WHO, 2007). In particular, there was little concordance of judgements when the high graded causality assignments of suspected kava hepatotoxicity proposed by the regulatory agency (BfArM, 2005) were further analyzed and found to be at best low graded by the WHO (2007). Among the 18 regulatory cases the regulatory agency primarily attributed causalities for kava as certain ($n = 1$), probable ($n = 11$), and possible/probable ($n = 2$) (BfArM, 2002, 2005); upon reassessment the WHO report coded a probable causality as the highest grade for only one single case from Germany (WHO, 2007). This discrepancy suggests specific problems of the two assessor groups of experts with their diverging causality results obtained with identical cases and methods. The problems may be due to the lack of a item by item presentation of the required features of the WHO scale by the regulatory agency when assessing the kava cases under consideration (BfArM, 2005); this regulatory approach does not support the concept of data transparency to allow health institutes and additional regulatory agencies as well as scientists an own substantiated assessment. Others have generally criticized the WHO scale on the basis of subjectivity and imprecision once it is mainly based on expert clinical judgements (Macedo et al., 2003). Overall, the impression prevails that lack of transparency and poor data quality of cases combined with the WHO scale as an inappropriate causality tool may create additional concern on scientific grounds. Therefore, the WHO scale appears to have no place for cases of liver injury to be assessed regarding causality.

5. Naranjo scale

The liver unspecific Naranjo scale (Naranjo et al., 1981b) with its various items (Table 3) has been praised and was used by the German regulatory agency to assess its cases of suspected kava hepatotoxicity regarding causality for kava (BfArM, 2005). The use of the Naranjo scale for patients with assumed toxic liver disease has been criticized on general grounds (García-Cortés et al., 2008) and in reference to both drug-induced liver injury (Andrade et al., 2001, 2004; García-Cortés et al., 2003, 2004, 2008) and herbal hepatotoxicity (Liss and Lewis, 2009; Teschke et al., 2009b).

Table 3
Naranjo scale.

	Yes	No	Do not know
Items			
1. Are there previous conclusive reports on this reaction?	+1	0	0
2. Did the adverse event appear after the suspected drug was administered?	+2	−1	0
3. Did the adverse reaction improve when the drug was discontinued or a specific antagonist was administered?	+1	0	0
4. Did the adverse reaction reappear when the drug was readministered?	+2	−1	0
5. Are there alternative causes (other than the drug) that could on their own have caused the reaction?	−1	+2	0
6. Did the reaction reappear when a placebo was given?	−1	+1	0
7. Was the drug detected in the blood (or other fluids) in concentrations known to be toxic?	+1	0	0
8. Was the reaction more severe when the dose was Increased, or less severe when the dose was decreased?	+1	0	0
9. Did the patient have a similar reaction to the same or Increased, or less severe when the dose was decreased?	+1	0	0
10. Was the adverse event confirmed by any objective evidence?	+1	0	0

Causality ratings on the basis of total scores: definitive $\geqslant 9$, probable 5–8, possible 1–4, doubtful $\leqslant 0$. Details are derived from Naranjo et al. (1981a,b).

The Naranjo scale has been developed at a time when efforts were lacking to consider organ specificity (Naranjo et al., 1981b), as exemplified by its unspecific items (Table 3). This scale lacks specificity for hepatotoxicity, with its particular clinical and chronological characteristics, as the different criteria are not usually weighted to take into account differences between organs (Andrade et al., 2004). Therefore, its use appears limited by various factors as are missing definition of liver injury as adverse drug reaction; lack of clear time frames of latency period; undefined time frame for dechallenge; lacking definitions of risk factors; insufficient evaluation of alternative diagnoses; inappropriate assessment of comedicated drugs, herbal drugs, and dietary supplements containing also herbs; and lacking definition of a positive rechallenge test. Under these conditions, much is left to individual interpretation and open for discussion. Due to the multiplicity of possible organ systems involved in adverse drug reactions, the various items of the Naranjo scale were not based on any gold standards related to each individual organ; instead, for the original validation of the Naranjo scale published reports of adverse drug reactions were used, not considering organ specificity (Naranjo et al., 1981b). This led to major general discussions around this scale in terms of inaccuracy and its unspecificity for assessing adverse drug reactions which involve the liver as main target organ (García-Cortés et al., 2008; Liss and Lewis, 2009). The problems of the Naranjo scale may be underscored by the consideration that the weightings of criteria might differ among various adverse reactions to take into account the singularities of each therapeutic problem (García-Cortés et al., 2004, 2008). This certainly explains why in cases of nonliver events full agreement was achieved in only 35% between the assessors (Naranjo et al., 1981b).

The Naranjo algorithm was initially designed for evaluation of adverse drug reactions related to pharmacological actions of the drugs (Naranjo et al., 1981b); hence contains questions regarding such as drug concentrations and monitoring, dose relationship, and placebo response (Table 3), which are clearly not relevant to idiosyncratic drug-induced liver injury (García-Cortés et al., 2008). The Naranjo scale has other confusing questions regarding positive rechallenge and previous exposure or cross-reactivity, which are prone to different interpretations and responses. This applies also to questions regarding decreasing dose and dechallenge (Table 3). In addition, the last question of the Naranjo scale asks for a confirmation of the adverse drug reaction using objective evidence. The question can be answered differently depending on whether the clinical observer considers a compatible liver biopsy, a positive rechallenge, or just more or less increased ALT values as definite evidence of hepatotoxicity. Moreover, even definite criteria such as positive rechallenge showed disagreements among the answers of the observers (García-Cortés et al., 2008).

Going into other details of the Naranjo scale with the questionnaire as listed item by item (Table 3), a possible causality category requires 1–4 points and is easily achieved: two points may theoretically be awarded just by the fact that the incriminated product was used prior to the assumed liver injury, even in the absence of any additional parameter(s) commonly required for a sound causality assessment. Under these circumstances, lack of liver values may award 0 points provided the adverse reaction was not confirmed by any objective evidence such as increased ALT values; yet a possible causality grading may nevertheless be maintained although hepatotoxicity was in no way established. Conversely, when liver values are increased (resulting in 1 point) and the adverse event did not appear after the use of the incriminated product (yielding 0 points), causality is theoretically still considered possible with 1 point. These few examples illustrate the flaws of the Naranjo scale which provides a possible causality category under almost any condition(s), even when there is lack of a temporal association or liver disease was not firmly established. This explains causalities of possible categories for herbs found with the Naranjo scale (Liss and Lewis, 2009; Teschke et al., 2009b).

Additional details of the Naranjo scale have been provided by thorough analyses of cases of drug-induced liver injury (García-Cortés et al., 2008). In the latter study with drug-induced liver injury, an overall agreement between assessors was achieved in 45% and thereby no better than the 50% agreement found if the observers had evaluated general adverse reactions without using a diagnostic scale (Naranjo et al., 1981a). In this particular context the conclusion was reached that the Naranjo scale does not add consistency or objectivity to the causality assessment of drug-induced liver injury cases assessed on clinical grounds (García-Cortés et al., 2008). In addition, the items about confirmation of the adverse event by an objective evidence, effect of rechallenge, and exclusion of other causes showed different levels of disagreement of 15%, 37%, and 71%, respectively (García-Cortés et al., 2008). The latter figure is particularly high and shows that disagreements between observers in the application of the Naranjo scale were related to the use of clinical judgement in evaluating alternative etiological explanations. These included nondrug and drug-related causes, as a standardized methodology is not provided. This deficiency has already been highlighted in the original description of the Naranjo scale (Naranjo et al., 1981b; Lanctot and Naranjo, 1995; García-Cortés et al., 2008). It has also been pointed out that the Naranjo scale, if applied to cases of drug-induced liver imjury, has a low sensitivity (54%) and poor negative predictive value (29%) and showed a limited capability to distinguish between adjacent categories and probability (García-Cortés et al., 2008). Thus, the Naranjo scale lacks validity and reproducibility in the attribution of causality in hepatotoxicity.

The use of the Naranjo scale in cases of suspected herbal hepatotoxicity has been criticized on various grounds and is discouraged in this particular disease setting (Liss and Lewis, 2009; Teschke et al., 2009b). The Naranjo scale was considered too insensitive to be reliably used in assessing primarily suspected herbal liver injury, as illustrated by a regulatory series on green tea

extracts, in which all 34 instances of purported hepatic injury were judged as being at least possibly related by virtue of the patient simply having taken the suspected agent (Liss and Lewis, 2009). It has also been emphasized that the Naranjo scale is unfortunately not specific to toxic liver injury and often allows for a score of a possible relationship, even in the absence of essential data. Concomitantly, concern has been expressed that a more discriminating judgement would have been expected from the regulatory agency. Similar problems emerged when the Naranjo scale was used for regulatory cases with primarily suspected herbal hepatotoxicity by black cohosh, since various unrelated causes have been established subsequent to thorough analyses (Teschke et al., 2009a,b). Missed and therefore not made diagnoses included genuine autoimmune hepatitis; alcoholic or cardiac hepatopathy; hepatotoxicity induced by Interferon and Fluoxetine; unspecific marginally or moderately increased serum activities of alanine aminotransferase or gamma-glutamyltranspeptidase without clinical relevance and not meeting criteria of liver injury; preexisting liver diseases; and rosuvastin-induced rhabdomyolysis. In other cases, data quality was so poor that assessment was not feasible. All these confounding factors were not recognized by the Naranjo scale; it therefore appears that the Naranjo scale is an invalid diagnostic approach for causality assessment in patients with primarily assumed herbal hepatotoxicity, justifying the critical comments.

For causality assessment of primarily suspected kava hepatotoxicity, the regulatory agency used the Naranjo scale for its cases (BfArM, 2005). However, this recent regulatory evaluation lacked any description of a final score for each proposed case, and there was also no item by item presentation of the cases required for reasons of transparency. Some cases received intermediate causality categories such as possible/probable, although the Naranjo scale does not provide for such a category. Any evaluation by the Naranjo scale usually results in generating full numbers of scores for each individual patient rather than in partial ones, and full numbers signify clearly the corresponding causality classification (Table 3). Due to these shortcomings and uncertainties a reassessment of the cases with the Naranjo scale is not feasible. Therefore, data of kava cases obtained with the Naranjo scale have to be taken with caution, and this applies also to other cases of suspected liver injury under regulatory consideration.

Taken together, it appears that the Naranjo scale is an invalid diagnostic approach for causality assessment in patients with assumed liver injury caused not only by kava but also by other herbs and also by conventional drugs. Despite the shortcomings in connection with liver injury, the Naranjo scale may have its merits for assessments of adverse drug reactions unrelated to the liver.

6. CIOMS scale and main-test

It was unexpected that none of the published regulatory evaluations of primarily assumed kava hepatotoxicity refers to liver specific causality assessment methods considered to be used (BfArM, 2002, 2005). In particular, the structured, quantitative and hepatotoxicity specific causality assessment method of CIOMS (Council for International Organizations of Medical Sciences) (Danan and Bénichou, 1993; Bénichou et al., 1993) was neither discussed nor mentioned in this context by the regulatory agency, it has simply been ignored (BfArM, 2002, 2005). This became a matter of concern and raised serious doubts as to what extent the regulatory agency was in a position to establish causality using specific and well accepted tools (Teschke et al., 2008a; Teschke and Wolff, 2009). Of note, EMA used the CIOMS scale to evaluate cases of primarily suspected herbal hepatotoxicity (Teschke et al., 2009a,b).

The CIOMS scale (Danan and Bénichou, 1993; Bénichou et al., 1993) and the main-test as its updated scale (Table 4) (Teschke et al., 2008b) represent well validated structured, quantitative and hepatotoxicity specific causality assessment methods and were employed for evaluation of suspected kava hepatotoxicity in a total of 31 cases (Teschke et al., 2008a; Teschke, 2010a). Causality for kava ± comedication was finally established in 14 of these patients with causality categories of highly probable, probable, or possible. The rate of both missed and alternative diagnoses was high in contrast to the regulatory evaluation (BfArM, 2005). Using the CIOMS scale or the main-test, item by item may be evaluated and published (Table 4); this approach facilitates both transparency and subsequent reevaluation through the scientific community. Key elements required for causality assessment of drug and herbal hepatotoxicity include various items such as temporal association; dechallenge with exact course of liver enzymes; exclusion of hepatitis A, B, and C, CMV, EBV, HSV, VZV, biliary obstruction, and cardiac hepatopathy; and comedication (Danan and Bénichou, 1993; Teschke et al., 2008b). For the main-test, all items of the CIOMS scale have been transferred and used (Table 4); for reasons of precision and actualization, a diagnostic update regarding serology and PCR was required and made for infections by hepatitis and hepatotropic viruses, and the existing item of hepatobiliary sonography was supplemented by the routinely applied colour Doppler sonography of the liver vessels (Teschke et al., 2008b). These basic features of the main-test (Table 4) are not adequately considered by the ad hoc causality assessment method (Table 1), the WHO scale (Table 2), and the Naranjo scale (Table 3). For causality assessments in cases of suspected liver injury, additional methods such as the pre-test for a quick evaluation and the post-test for exclusion of various differential diagnoses are available and useful (Teschke et al., 2008b).

Although in need for further refinement, the CIOMS scale and/or the main-test have been used for quantitative causality assessments of suspected hepatotoxicity by prescription drugs and herbal medications in a variety of studies such as epidemiological studies, clinical trials, case reports, case series, regulatory analyses, and genotyping studies (Kaplowitz, 2001; Andrade et al., 2005; Rochon et al., 2008; García-Cortés et al., 2008; Teschke and Bahre, 2009; Teschke et al., 2008a, 2009a,b; Rockey et al., 2010). Overall, the CIOMS scale is judged a reliable and reproducible tool, providing an optimum level of objectivity (García-Cortés et al., 2008). It is of considerable clinical value in assessing complex cases of patients and in research settings. Further, the scale is useful in routine clinical practice to recall the parameters that need to be systematically addressed in cases of suspected hepatotoxicity so that clinical judgement can be improved and become more consistent. Supportive evidence is also lacking that the CIOMS scale could be replaced by liver unspecific methods such as the Naranjo scale (García-Cortés et al., 2008). A simplified version of the CIOMS scale in form of the Clinical Diagnostic Scale or the MV scale (Maria and Victorino, 1997) has previously been discussed and applied in patients with drug-induced liver injury (Aithal et al., 2000; Lee, 2000; Lucena et al., 2001; Kaplowitz, 2001; Teschke et al., 2008b); it was recently used to assess causality in the regulatory cases of primarily suspected kava hepatotoxicity, but performance was poor compared to the original CIOMS scale (Teschke et al., 2010). Therefore, structured hepatotoxicity specific causality assessment methods such as the CIOMS and the main-test are clearly the preferred tools for causality assessment of liver injury by drugs and herbs including kava, and this recommendation should also be acceptable for regulatory agencies and health institutes.

7. Future aspects

Regulatory agencies and health institutes have commonly a good reputation in the area of pharmocovigilance with special expertise in adverse reactions elicited by conventional drugs and herbs. Issues emerged, however, when cases of liver injury had to

Table 4
Main-test (hepatocellular injury).

Hepatocellular injury	Score	Patient
1. Time to onset from the beginning of the drug		
• 5–90 days (rechallenge: 1–15 days)	+2	
• <5 or >90 days (rechallenge: >15 days)	+1	
2. Time to onset from cessation of the drug		
• ⩽15 days (except for slowly metabolized drugs: >15 days)	+1	
3. Course of ALT after cessation of the drug		
Difference between peak of ALT and upper limit of normal range		
• Decrease ⩾ 50% within 8 days	+3	
• Decrease ⩾ 50% within 30 days	+2	
• No information	0	
• Decrease ⩾ 50% after the 30th day	0	
• Decrease < 50% after the 30th day or recurrent increase	−2	
4. Risk factor ethanol		
• Yes	+1	
• No	0	
5. Risk factor age		
• ⩾55 years	+1	
• <55 years	0	
6. Concomitant drug(s)		
• None or no information	0	
• Concomitant drug with incompatible time to onset	0	
• Concomitant drug with compatible or suggestive time to onset	−1	
• Concomitant drug known as hepatotoxin and with compatible or suggestive time to onset	−2	
• Concomitant drug with evidence for its role in this case (positive rechallenge or validated test)	−3	
7. Search for non drug causes		
Group I (6 causes)		
• Anti-HAV-IgM		
• Anti-HBc-IgM/HBV-DNA		
• Anti-HCV-IgM/HCV-RNA		
• Hepatobiliary sonography/colour Doppler sonography of liver vessels		
• Alcoholism (AST/ALT ⩾ 2)		
• Acute recent hypotension history (particularly if underlying heart disease)		
Group II		
• Complications of underlying disease(s)		
• Infection suggested by PCR and titer change for		
CMV (Anti-CMV-IgM/IgG)		
EBV (Anti-EBV-IgM/IgG)		
HSV (Anti-HSV-IgM/IgG)		
VZV (Anti-VZV-IgM/IgG)		
Evaluation of group I and II		
• All causes-groups I and II – reasonably ruled out	+2	
• The 6 causes of group I ruled out	+1	
• 5 or 4 causes of group I ruled out	0	
• Less than 4 causes of group I ruled out	−2	
• Non drug cause highly probable	−3	
8. Previous information on hepatotoxicity of the drug		
• Reaction labelled in the product characteristics	+2	
• Reaction published but unlabelled	+1	
• Reaction unknown	0	
9. Response to readministration		
• Doubling of ALT with the drug alone	+3	
• Doubling of ALT with the drugs already given at the time of 1st reaction	+1	
• Increase of ALT but less than *N* in the same conditions as for the first administration	−2	
• Other situations	0	
Total points for patient:		

The term drug is used for synthetic drugs, herbal drugs, and dietary supplements including herbal ones. ALT: alanine aminotransferase, AST: asparate aminotransferase, HAV: hepatitis A virus, HBc: hepatitis B core, HBV: hepatitis B virus, HCV: hepatitis C virus, CMV: cytomegalovirus, EBV: Epstein Barr virus, HSV: herpes simplex virus, VZV: varicella zoster virus. Total points/causality: ⩽0 = excluded; 1–2 = unlikely; 3–5 = possible; 6–8 = probable; and >8 = highly probable. Data are derived from Teschke et al. (2008b).

be assessed regarding causality. To overcome these problems, new and stringent approaches will be required for causality assessment of suspected cases of liver injury by herbs and drugs. In particular, there is an urgent need to use liver specific causality assessment methods such as the CIOMS scale and the main-test instead of liver unspecific approaches like ad hoc evaluation methods, the WHO scale, or the Naranjo scale. Causality assessment is facilitated by installation of a sophisticated data collection system with the intention to receive more complete data of the cases reported by treating physicians to regulatory agencies.

8. Conclusions

In conclusion, the analysis shows that liver specific causality assessment methods such as the CIOMS scale and the main-test as its update are the preferred tools for the evaluation of primarily suspected kava hepatotoxicity, whereas the use liver unspecific methods such as the ad hoc causality approach, the WHO scale or the Naranjo scale are considered obsolete under these conditions. Regulatory agencies and health institutes are therefore well advised to use in future the appropriate liver specific causality assessment tools for cases of primarily suspected herbal hepatotoxicity. Transparency is urgently needed regarding data details of each case with suspected liver injury by conventional drugs and herbal dietary herbal supplements containing also kava to improve pharmacovigilance and patient safety. Item by item presentation of well documented cases should be the primary goal in the area of clinical and regulatory drug and herbal hepatotoxicity.

Conflicts of interest

None declared.

References

Aithal, G.P., Rawlins, M.D., Day, C.P., 1999. Accuracy of hepatic adverse drug reactions reported in one English health region. Brit. Med. J. 319, 1541.

Aithal, G.P., Rawlins, M.D., Day, C.P., 2000. Clinical diagnostic scale: a useful tool in the evaluation of suspected hepatotoxic adverse drug reactions. J. Hepatol. 33, 949–952.

Andrade, R.J., Camargo, R., Lucena, M.I., González-Grande, R., 2004. Causality assessment in drug-induced hepatotoxicity. Expert Opin. Drug Saf. 3, 329–344.

Andrade, R.J., Guilarte, J., Salmerón, F.J., Lucena, M.I., Bellot, V., 2001. Benzylpenicillin-induced prolonged cholestasis. Ann. Pharmacother. 35, 783–784.

Andrade, R.J., Lucena, M.I., Fernández, M.C., Pelaez, G., Pachkoria, K., García-Ruiz, E., García-Muñoz, B., Gonzalez-Grande, R., Pizarro, A., Durán, J.A., Jiménez, M., Rodrigo, L., Romero-Gomez, M., Navarro, J.M., Planas, R., Costa, J., Borras, A., Soler, A., Salmerón, J., Martin-Vivaldi, R.Spanish Group for the Study of Drug-induced Liver Disease, 2005. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 129, 512–521.

Andrade, R.J., Lucena, M.I., Kaplowitz, N., García-Muñoz, B., Borraz, Y., Pachkoria, K., García-Cortés, M., Fernández, M.C., Pelaez, G., Rodrigo, L., Durán, J.A., Costa, J., Planas, R., Barriocanal, A., Guaner, C., Romero-Gomez, M., Muṇoz-Yagüe, T., Salmerón, J., Hidalgo, R., 2006. Outcome of acute idiosyncratic drug-induced liver injury: long term follow-up in a hepatotoxicity registry. Hepatology 44, 1581–1588.

Bénichou, C., Danan, G., Flahault, A., 1993. Causality assessment of adverse reactions to drugs – II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. J. Clin. Epidemiol. 46, 1331–1336.

BfArM (Bundesinstitut für Arzneimittel und Medizinprodukte, Bonn. Federal Institute for Drugs and Medicinal Products in Germany), 2002. Rejection of Drug Risks, Step II. As Related to: Kava–Kava (*Piper methysticum*)-containing, and Kavain-containing Drugs, Including Homeopathic Preparations with a Final Concentration up to, and Including D4, June 14, 2002. <http://www.spc.int/cis/documents/02_0714_BfArM_Kava_Removal.pdf> (accessed 20.07.10).

BfArM (Bundesinstitut für Arzneimittel und Medizinprodukte, Bonn. Federal Institute for Drugs and Medicinal Products in Germany), 2005. Kava–Kava, May 12, 2005. <http://www.bfarm.de/DE/Pharmakovigilanz/risikoinfo/functions/2005/risikoinfo-2005-node.html> (accessed 20.07.10).

Borrelli, F., Ernst, E., 2008. Black cohosh (*Cimicifuga racemosa*): a systematic review of adverse events. Am. J. Obstet. Gynecol. 199, 455–466.

Danan, G., Bénichou, C., 1993. Causality assessment of adverse reactions to drugs – I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. J. Clin. Epidemiol. 46, 1323–1330.

Denham, A., McIntyre, M., Whitehouse, J., 2002. Kava – the unfolding story: report on a work-in-progress. J. Altern. Complement. Med. 8, 237–263.

García-Cortés, M., Lucena, M.I., Andrade, R.J., Romero-Gomez, M., Fernández, M.C., 2003. Lansoprazole-induced hepatic dysfunction. Ann. Pharmacother. 37, 1731.

García-Cortés, M., Lucena, M.I., Andrade, R.J., Camargo, R., Alcántara, R., 2004. Is the Naranjo probability scale accurate enough to ascertain causality in drug-induced hepatotoxicity? Ann. Pharmacother. 38, 1540–1541.

García-Cortés, M., Lucena, M.I., Pachkoria, K., Borraz, Y., Hidalgo, R., Andrade, R.J., 2008. Evaluation of Naranjo Adverse Drug Reactions Probability Scale in causality assessment of drug-induced liver injury. Aliment. Pharmacol. Ther. 27, 780–789.

Gunawan, B., Kaplowitz, N., 2004. Clinical perspectives on xenobiotic-induced hepatotoxicity. Drug Metab. Rev. 36, 301–312.

Health Canada, 2010. Black cohosh products and liver toxicity: update. Canadian Adverse React. Newslett. 20, 1–3. <http://www.hc-sc.gc.ca/dhp-mps/medeff/bulletin/carn-bcei_v20n1-eng.php#a1t1> (accessed 20.07.10).

Kaplowitz, N., 2001. Causality assessment versus guilt-by-association in drug hepatotoxicity. Hepatology 33, 308–310.

Lanctot, K.L., Naranjo, C.A., 1995. Comparison of the Bayesian approach and a simple algorithm for assessment of adverse drug events. Clin. Pharmacol. Ther. 58, 692–698.

Lee, W.M., 2000. Assessing causality in drug-induced liver injury. J. Hepatol. 33, 1003–1005.

Lee, W.M., 2003. Drug-induced hepatotoxicity. New Engl. J. Med. 349, 474–485.

Liss, G., Lewis, J.H., 2009. Drug-induced liver injury: what was new in 2008? Expert Opin. Drug Metab. Toxicol. 5, 843–860.

Lucena, M.I., Camargo, R., Andrade, R.J., Perez-Sanchez, C.J., Cuesta, F.S.D.L., 2001. Comparison of two clinical scales for causality assessment in hepatotoxicity. Hepatology 33, 23–30.

Macedo, A.F., Marques, F.B., Ribeiro, C.F., Teixeira, F., 2003. Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel, according to different levels of imputability. J. Clin. Pharm. Ther. 28, 137–143.

Maddrey, W.C., 2005. Drug-induced hepatotoxicity 2005. J. Clin. Gastroenterol. 39, S83–S89.

Maria, V.A., Victorino, R.M., 1997. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. Hepatology 26, 664–669.

Naranjo, C.A., Busto, U., Abel, J.G., Sellers, E.M., 1981a. Empiric delineation of the probability spectrum of adverse drug reactions. Clin. Pharmacol. Ther. 29, 267–268.

Naranjo, C.A., Busto, U., Sellers, E.M., Sandor, P., Ruiz, I., Roberts, E.A., Janecek, E., Domecq, C., Greenblatt, D.J., 1981b. A method for estimating the probability of adverse drug reactions. Clin. Pharmacol. Ther. 30, 239–245.

Navarro, V.J., 2009. Herbal and dietary supplement hepatotoxicity. Semin. Liver Dis. 29, 373–382.

Rochon, J., Protiva, P., Seeff, L.B., Fontana, R.J., Liangpunsakul, S., Watkins, P.B., Davern, T., McHutchison, J.G., 2008. For the Drug-Induced Liver Injury Network (DILIN). Reliability of the Roussel Uclaf Causality Assessment Method for assessing causality in drug-induced liver injury. Hepatology 48, 1175–1183.

Rockey, D.C., Seeff, L.B., Rochon, J., Freston, J., Chalasani, N., Bonachini, M., Fontana, R.J., Hayashi, P.H., 2010. For the US Drug-induced Liver Injury Network. Causality assessment in drug-induced liver injury using a structured expert opinion process: comparison to the Roussel Uclaf Causality Assessment Method. Hepatology 51, 2117–2126.

Schmidt, M., Morgan, M., Bone, K., McMillan, J., 2005. Kava: a risk-benefit assessment. In: Mills, M., Bone, K. (Eds.), The Essential Guide to Herbal Safety. Elsevier Churchill Livingstone, St. Louis (Missouri), pp. 155–221.

Schulze, J., Raasch, W., Siegers, C.P., 2003. Toxicity of kava pyrones, drug safety and precautions – a case study. Phytomedicine 10 (Suppl. IV), 68–73.

Seeff, K.B., 2007. Herbal hepatotoxicity. Clin. Liver Dis. 11, 577–596.

Teschke, R., 2010a. Kava hepatotoxicity: a clinical review. Ann. Hepatol. 9, 251–265.

Teschke, R., 2010b. Kava hepatotoxicity: pathogenetic aspects and prospective considerations. Liver Int. 30, 1270–1278.

Teschke, R., Bahre, R., 2009. Severe hepatotoxicity by Indian Ayurvedic herbal products: a structured causality assessment. Ann. Hepatol. 8, 258–266.

Teschke, R., Bahre, R., Fuchs, J., Wolff, A., 2009a. Black cohosh hepatotoxicity: quantitative causality evaluation in nine suspected cases. Menopause 16, 956–965.

Teschke, R., Bahre, R., Genthner, A., Fuchs, J., Schmidt-Taenzer, W., Wolff, A., 2009b. Suspected black cohosh hepatotoxicity – challenges and pitfalls of causality assessment. Maturitas 63, 302–314.

Teschke, R., Bahre, R., Genthner, A., Fuchs, J., Schmidt-Taenzer, W., Wolf, A., 2009c. Suspected black cohosh hepatotoxicity – causality assessment versus safety signal. Quality versuch quantity. Maturitas 64, 141–142.

Teschke, R., Fuchs, J., Bahre, R., Genthner, A., Wolff, A., 2010. Kava hepatotoxicity: comparative study of two structured quantitative methods for causality assessment. J. Clin. Pharm. Ther. 34. doi:10.1111/j.1365-2710.2009.01131.x.

Teschke, R., Gaus, W., Loew, D., 2003. Kava extracts: safety and risks including rare hepatotoxicity. Phytomedicine 10, 440–446.

Teschke, R., Schwarzenboeck, A., Hennermann, K.H., 2008a. Kava hepatotoxicity: a clinical survey and critical analysis of 26 suspected cases. Eur. J. Gastroenterol. Hepatol. 20, 1182–1193.

Teschke, R., Schwarzenboeck, A., Hennermann, K.H., 2008b. Causality assessment in hepatotoxicity by drugs and dietary supplements. Brit. J. Clin. Pharmacol. 66, 758–766.

Teschke, R., Wolff, A., 2009. Kava hepatotoxicity: regulatory data selection and causality assessment. Dig. Liver Dis. 41, 891–901.

WHO, 2000. Causality Assessment of Suspected Adverse Reactions. WHO Collaboratoring Centre for International Drug Monitoring (Uppsala Monitoring Centre, UMC). Database, 2000. <http://www.who-umc.org/DynPage.aspx?id=22682> (accessed 20.07.10).

WHO (World Health Organization), 2007. Assessments of the Risk of Hepatotoxicity with Kava Products. WHO Document Production Services, Geneva, Switzerland.

Zimmerman, H.J., 1999. Hepatotoxicity. Lippincott Williams & Wilkins, Philadelphia.