

Evaluation of Naranjo Adverse Drug Reactions Probability Scale in causality assessment of drug-induced liver injury

M. GARCÍA-CORTÉS*, M. I. LUCENA†, K. PACHKORIA†, Y. BORRAZ†, R. HIDALGO‡ & R. J. ANDRADE* ON BEHALF OF THE SPANISH GROUP FOR THE STUDY OF DRUG-INDUCED LIVER DISEASE (GRUPO DE ESTUDIO PARA LAS HEPATOPATÍAS ASOCIADAS A MEDICAMENTOS, GEHAM)

*Unidad de Hepatología, Servicio de Aparato Digestivo, Hospital Universitario Virgen de la Victoria, Facultad de Medicina, Campus Universitario de Teatinos sn, Málaga; †Servicio de Farmacología Clínica, Grupo de Estudio para las Hepatopatías Asociadas a Medicamentos, The Spanish Group for the Study of Drug-Induced Liver Disease (GEHAM), Co-ordinating Centre, Hospital Universitario Virgen de la Victoria, Facultad de Medicina, Campus Universitario de Teatinos s/n, Málaga; ‡Centro de Cálculo, Universidad de Málaga, Málaga, Spain

Correspondence to:
Prof R. J. Andrade, Departamento de Medicina, Facultad de Medicina, Boulevard Louis Pasteur, 32, Campus de Teatinos s/n, 29071 Málaga, Spain.
E-mail: andrade@uma.es

Publication data

Submitted 10 January 2008
First decision 10 February 2008
Resubmitted 12 February 2008
Accepted 14 February 2008
Epub Online Accepted 18 February 2008

SUMMARY

Background

Causality assessment in hepatotoxicity is challenging. The current standard liver-specific Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method scale is complex and difficult to implement in daily practice. The Naranjo Adverse Drug Reactions Probability Scale is a simple and widely used nonspecific scale, which has not been specifically evaluated in drug-induced liver injury.

Aim

To compare the Naranjo method with the standard liver-specific Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method scale in evaluating the accuracy and reproducibility of Naranjo Adverse Drug Reactions Probability Scale in the diagnosis of hepatotoxicity.

Methods

Two hundred and twenty-five cases of suspected hepatotoxicity submitted to a national registry were evaluated by two independent observers and assessed for between-observer and between-scale differences using percentages of agreement and the weighted kappa (κ_w) test.

Results

A total of 249 ratings were generated. Between-observer agreement was 45% with a κ_w value of 0.17 for the Naranjo Adverse Drug Reactions Probability Scale, while there was a higher agreement when using the Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method scale (72%, κ_w : 0.71). Concordance between the two scales was 24% (κ_w : 0.15). The Naranjo Adverse Drug Reactions Probability Scale had low sensitivity (54%) and poor negative predictive value (29%) and showed a limited capability to distinguish between adjacent categories of probability.

Conclusion

The Naranjo scale lacks validity and reproducibility in the attribution of causality in hepatotoxicity.

Aliment Pharmacol Ther 27, 780–789

INTRODUCTION

Drug-induced liver injury (DILI) is a leading health problem especially in a globally expanding commercialization of new drugs and the increasing exposure of patients to new compounds. This is expected to increase because of the number of drugs being prescribed and because of the current tendency towards pharmacologically active complementary and alternative medicines. As a consequence, DILI ranks as the first cause of acute liver failure in the US and Scandinavian countries and is the main reason for postcommercialization decisions by regulatory bodies and withdrawal of drugs from the market.¹⁻⁴ The real incidence of DILI remains unknown and difficult to estimate effectively.⁵ Thus, in a community-based prospective survey conducted in a rural area of France, the overall incidence of DILI was estimated as 13.9 cases per 100 000 people per year. This was 16 times higher than the number of cases reported to the French drug monitoring system.⁶ This discrepancy highlights the importance of increasing the awareness of adverse hepatic reactions among prescribers of drugs as well as of integrating into routine clinical practice the necessary tools to increase the degree of certainty in diagnosing DILI.

There are no reliable markers for the diagnosis of DILI and, currently, the reliance is on circumstantial evidence and on the exclusion of other causes of liver injury.⁷ To improve consistency, accuracy and objectiveness, several different clinical instruments for the causality assessment of adverse drug reactions have been proposed.⁸ The Naranjo Adverse Drug Reactions Probability Scale (NADRPS) was developed⁹ for the assessment of adverse drug reactions. Despite its simplicity and wide applicability, the NADRPS is yet to be systematically evaluated with respect to DILI. More recently, a specific scale for the causality assessment of DILI was derived from an international consensus meeting of experts,^{10, 11} and was termed the Council for International Organizations of Medical Sciences or Roussel Uclaf Causality Assessment Method (CIOMS or RUCAM) scale. The CIOMS algorithm has been shown to be more reliable and reproducible than other scales in the evaluation of DILI¹² and, currently, it is the standard instrument for causality assessment of hepatotoxicity. However, it is complex and difficult to implement in daily practice.

Some journals require the application of the NADRPS for the assessment of any adverse drugs reaction (including hepatic reactions) as a prerequisite for con-

sideration of a manuscript for publication, but there have been concerns regarding the accuracy and reproducibility of NADRPS in DILI cases.¹³⁻¹⁵ Hence, this study was conducted to evaluate the performance of the NADRPS in comparison with the CIOMS standard scale, in a large cohort of cases of hepatotoxicity collated in a national registry.

METHODS

Case selection

The last 225 consecutive patients with suspected drug-induced liver disease submitted to the Spanish Registry of Hepatotoxicity were reviewed. The operational structure of the registry, data recording and case ascertainment has been reported previously.¹⁶ In addition, several cases published in peer-reviewed journals as case report or case series have been included in this evaluation.¹⁷⁻²⁴

In brief, clinicians identified and recorded data from hospitalized patients as well as from outpatients with suspected DILI, and submitted this information to the coordinating centre using a structured reporting form, which includes demographic data, a detailed clinical history including underlying diseases and treatment, the presence of risk factors such as alcohol intake or pregnancy, temporal relationship between the prescription of the suspected drug and the onset of liver injury, and between the withdrawal of the drug and the course of the reaction, exclusion of alternative causes of liver damage and the outcome of the reaction. Serology and specific biochemistry to rule out viral hepatitis, autoimmune and metabolic liver disorders, as well as appropriate imaging tests to exclude biliary disease were made. A liver biopsy was considered compatible when any of the following features were involved: centrilobular predominance of the lesion, eosinophilic infiltrates, granulomas, bile duct injury and microvesicular steatosis.

Cases with suspected DILI reported to the registry were initially made by the clinical judgement of doctor-in-charge and, subsequently, by the evaluation of three experts based on the following criteria: (i) an appropriate temporal relationship between the intake of the drug and the onset of the event, (ii) the course of the reaction following the withdrawal of the drug, (iii) exclusion of other causes of liver disease, (iv) relapse following re-exposure when applicable and

(v) previous reports of the adverse reaction. This last category was confirmed by Product Summary Characteristics of each drug, by data in Pharmacopoeia Compendia such as Martindale's, and from Internet sources such as The National Library of Medicine (PubMed).²⁵

The pattern of liver damage is classified according to the International Consensus Meeting criteria.²⁶ Alternatively, liver damage was based on liver biopsy findings, when available. Cases were classified as 'intrinsic' or 'immunoallergic' when features of hypersensitivity were present. In the remaining cases, the mechanism was assumed to be idiosyncratic metabolic. Chronic evolution of the disease was defined as persistent abnormality of the laboratory liver tests >3 months following cessation of the culprit drug in the case of hepatocellular pattern of damage; or >6 months for cholestatic/mixed type of liver injury. Drugs were classified according to the Anatomical Therapeutic Classification recommended by the World Health Organization.²⁷

The study protocol was approved by the ethics committee of the coordinating centre at the 'Virgen de La Victoria' University Hospital of Málaga.

Scoring system

Two independent observers assessed each case using two different scales, the NADRPS and the CIOMS/RUCAM scale (Table 1). These methods are based on questionnaires that seek an operational definition of adverse drug reaction using a series of preset questions with different weightings depending on the response to each question.

The NADRPS involves 10 'yes', 'no' or 'unknown or non-applicable' questions. The adverse drug reaction are assigned to a probability category on the basis of the total score as 'definite' = ≥ 9 , 'likely' = 5–8, 'possible' = 1–4, 'unlikely' = ≤ 0 .

The CIOMS/RUCAM scale provides a scoring system for seven axes of decision strategy, where the responses correspond to weighted numerical values that are summed to give a total score. The categories of suspicion are 'definite or highly probable' (score >8), 'probable' (score 6–8), 'possible' (score 3–5), 'unlikely' (score 1–2) and 'excluded' (score ≤ 0).

To establish a common range of categories to enable direct comparisons, in the CIOMS/RUCAM method, the categories 'unlikely' and 'excluded' were grouped and renamed as 'unlikely'.

Data management and statistical analysis

Statistical analyses were performed with chi-squared contingency tables. Concordance between the numerical scores yielded by the two independent observers in the evaluation of each scale was estimated by the linear correlation coefficient (r). A P -value of <0.05 was considered statistically significant. The percentage of agreement and the weighted kappa (κ_w) were obtained. The result of this statistic tool has a maximum value of 1 when agreement between the observers is perfect, whereas a value of 0 indicates no agreement better than chance.^{28, 29}

Data were managed using the statistical package for the social sciences (SPSS; version 14.0; SPSS Inc., Chicago, IL, USA).

RESULTS

The NADRPS and the CIOMS/RUCAM scales were applied in a total of 225 cases, of which 193 patients met all the inclusion criteria for DILI and 32 cases were added as a control group because an alternative cause was identified at a later date. A total of 249 ratings were generated because of more than one culprit drug being potentially implicated in 24 cases. Table 2 summarizes the demographic characteristics, and laboratory findings of the individuals with the diagnosis of DILI according to main therapeutic classes involved. Mean age was 54 years (range: 13–87) and there was no difference in gender distribution (53% males). The predominant pattern of liver injury was hepatocellular (59% of cases). Hypersensitivity features were present in 21% of the patients. Of the group of DILI patients, eight had underlying hepatitis C virus infection, one was co-infected with the human immunodeficiency virus and three cases were hepatitis B virus inactive carriers. One hundred and nine patients (56%) needed hospitalization. Chronic outcome was evident in six cases. Death or liver transplantation occurred in 11 (6%) cases. The pharmacological groups of drugs most commonly implicated in DILI were anti-infective (30%), followed by nervous system drugs (19%), musculo-skeletal (13%) and cardiovascular drugs (10%). Amoxicillin-clavulanic acid appeared to be the leading individual drug combination leading to DILI (34 cases; 16%). 41 cases had a liver biopsy, the reason being cholestatic hepatitis and hepatocellular necrosis the predominant patterns.

Table 1. Scores for individual axes of the Naranjo Adverse Drug Reactions Probability Scale and Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method scale

Naranjo scale Axis	Numerical score	CIOMS/RUCAM scale	Type of liver injury				Score
			Hepatocellular		Cholestatic/mixed		
		Time of onset of the event:	First treatment	Second treatment	First treatment	Second treatment	
Previous reports on the reaction	0–1	From drug intake until reaction onset	5–90 days	1–15 days	5–90 days	1–90 days	+2
Temporal illegibility in the onset of the reaction	–1–2		<5 or >90 days	>15 days	<5 or >90 days	>90 days	+1
Improvement after drug withdrawal	0–1	From drug withdrawal until reaction onset	< or =15 days	< or =15 days	< or =30 days	< or =30 days	+1
Positive re-challenge	–1–2	Course of the reaction	>50% improvement 8 days				+3
Exclusion of alternative causes for the ADR	–1–2		>50% improvement 30 days		>50% improvement 180 days		+2
Placebo response	0–1		Lack of information or no improvement		Lack of information or no improvement		0
Drug concentration and monitoring	0–1		Worsening or <50% improvement 30 days				–1
Dose relationship	0–1	Risk Factors	Alcohol		Alcohol or pregnancy		+1
Previous exposure and cross reactivity	0–1		Age > or = 55 years		Age > or = 55 years		+1
Presence of any objective evidence	0–1	Concomitant therapy					–3–0
		Exclusion of nondrug-related causes					–3–2
Results: > or = 9 definitive; 5–8 probable; 1–4 possible; < or = 0 unlikely		Previous information on hepatotoxicity					0–+2
		Re-challenge					–2–+3
Results: >8 points definite; 6–8 points probable; 3–5 points possible; 1–2 points unlikely; <0 points excluded							

Agreement between-observers was achieved in 112 ratings (45%) when using the NADRPS; corresponding to a κ_w of 0.17 (95% CI: 0.11–0.24; Table 3). Conversely, agreement in interobservers was obtained in 179 ratings (72%) using the CIOMS/RUCAM scale (Table 4) with a κ_w score of 0.71 (95% CI: 0.65–0.78); furthermore, adding complete agreement and discrepancies of one level ('definite' vs. 'probable') raised the percentage agreement to 92% (229 of 249). The numerical scores generated by both observers with the NADRPS and the CIOMS/RUCAM scales were compared using the Pearson correlation coefficient (r). The coefficients obtained were 0.52 ($P < 0.01$) for NADRPS and 0.89 ($P < 0.01$) for the CIOMS method.

A detailed analysis of the scores given to the different criteria of NADRPS by both observers, showed high agreement for the items referred to previous reports (99%), temporal relationship (99%) and improvement after withdrawal of the suspected drug (99%). Interestingly, the questions related to placebo response, drug concentration and monitoring, dose

relationship and previous exposure or cross-reactivity were scored by none of the raters. Alternatively, the items about exclusion of other causes, effect of re-challenge and confirmation of the adverse event by an objective evidence showed different levels of disagreement, 71%, 37%, and 15% respectively.

Concordance between the NADRPS and the CIOMS/RUCAM methods when data from observer I was combined with that from observer II is shown in Table 5. Of the total number of ratings, complete agreement was observed in 122 ratings (24%) with a κ_w of 0.15 (95% CI: 0.12–0.19). Disagreements of one level were found in 313 ratings (63%). Adding complete agreement and minor discrepancies raised the percentage agreement to 49% (244 of 498 ratings).

Outcomes remained unchanged with respect to different patterns of liver damage or the presence of hypersensitivity features when scales were compared. As a positive re-challenge is considered definitive evidence of DILI, we evaluated the scores given by the two scales for the 14 cases (7%) in whom liver injury

Table 2. Main laboratory findings by drug class in patients with drug-induced liver injury

Drug class	N	Gender M/F	Mean (range) of values			
			Age (years)	Total bilirubin (mg/dL)	ALT (×ULN)	AP (×ULN)
Anti-ulcer drugs	5	4/1	44 (13–69)	14.1 (6.1–18.5)	29.5 (13.9–62.8)	1.7 (1.1–2.5)
Oral hypoglycaemic drugs	3	1/2	69 (67–74)	6.4 (0.4–18.4)	6.7 (1.0–18.0)	2.9 (0.9–3.8)
Antithrombotic agents	4	4/0	73 (56–81)	6.5 (2.4–10.4)	10.3 (3.5–19.7)	1.5 (0.9–1.8)
Antidysrhythmics	3	1/2	55 (46–71)	2.4 (1.2–3.5)	38.2 (1.9–73.8)	1.0 (0.9–1.0)
Antihypertensives	8	6/2	62 (38–79)	11.6 (0.6–28.1)	10.8 (2.0–44.4)	3.2 (0.5–5.8)
Lipid lowering drugs	10	6/4	54 (43–69)	8.0 (1.0–26.0)	16.9 (1.9–47.9)	2.5 (0.6–11.1)
Antibiotics*	18	13/5	58 (25–87)	10.5 (1.6–19.7)	34.5 (2.4–85.5)	3.7 (0.6–32.7)
Amoxicillin–clavulanate	34	18/16	57 (16–83)	9.1 (0.4–32.6)	12.9 (1.7–64.9)	2.3 (0.4–7.9)
Antituberculous drugs	10	6/4	57 (32–79)	5.5 (0.7–11.6)	38.1 (5.6–86.4)	1.1 (0.3–1.7)
Antifungal drugs	4	3/1	53 (38–68)	14.2 (6.0–17.5)	18.7 (4.0–53.2)	1.8 (0.8–2.5)
Antineoplastic/immunomodulating agents	13	7/6	57 (15–79)	10.3 (0.4–26.9)	22.2 (2.2–98.8)	2.2 (0.6–7.1)
Nonsteroidal anti-inflammatory drugs	26	18/8	58 (30–83)	7.3 (0.8–33.9)	9.6 (0.9–43.8)	3.3 (0.5–14.2)
Antigout preparations	3	2/1	74 (74–74)	7.1 (4.6–12.0)	23.2 (4.5–32.2)	2.8 (2.4–3.7)
Analgesics and antipyretics	10	6/4	41 (19–72)	3.9 (0.4–18.6)	11.0 (2.4–29.4)	1.3 (0.3–2.9)
Antiepileptics	8	1/7	58 (24–73)	13.9 (0.2–36.9)	16.9 (3.7–60.5)	5.8 (0.5–9.7)
Antipsychotics	2	0/2	47 (43–51)	0.3 (0.3–0.4)	2.9 (2.7–3.1)	1.3 (0.4–2.2)
Anxiolytics, hypnotics and sedatives	9	2/7	56 (39–79)	3.3 (0.3–11.6)	26.7 (1.6–88.0)	1.2 (0.5–1.8)
Antidepressants	15	3/12	46 (31–63)	5.9 (0.1–38.5)	21.9 (3.5–95.9)	1.4 (0.4–4.5)
Sex hormones	8	1/7	43 (21–79)	14.5 (0.6–37.0)	12.5 (0.6–57.3)	1.3 (0.6–2.4)
Herbal remedies	5	0/5	37 (26–52)	9.9 (0.5–16.6)	48.8 (3.2–83.5)	0.9 (0.5–1.7)
Others	19	12/7	44 (18–78)	8.2 (0.3–35.9)	14.7 (1.2–41.3)	1.7 (0.4–3.6)

AP, alkaline phosphatase; ALT, alanine aminotransferase; F, female; M, male.

* Antibiotics except amoxicillin–clavulanic acid.

Values are expressed as multiples of the upper limit of normal (×ULN).

Table 3. Agreement between-observers applying the Naranjo scale

Observer I	Observer II				
	Definite	Probable	Possible	Unlikely	Total
Definite	0	1			1
Probable		50	3		53
Possible		126	60	3	189
Unlikely			4	2	6
Total	0	177	67	5	249

45% agreement, weighted kappa of 0.17.

Bold values highlight the actual number of cases in which full agreement was found between observers when applying the Naranjo scale.

Table 4. Agreement between-observers applying the Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method scale

Observer I	Observer II				
	Definite	Probable	Possible	Unlikely	Total
Definite	68	21			89
Probable	29	77	9		115
Possible		10	3		13
Unlikely			1	31	32
Total	97	108	13	31	249

72% agreement, weighted kappa of 0.71.

Bold values highlight the actual number of cases in which full agreement was found between observers when applying the CIOMS/RUCAM scale.

Table 5. Concordance between the Naranjo and the CIOMS/RUCAM scale of all cases scored by both observers

CIOMS/RUCAM	Naranjo				Total
	Unlikely	Possible	Probable	Definite	
Unlikely	10	53			63
Possible	1	14	11		26
Probable		126	97		223
Definite		63	122	1	186
Total	11	256	230	1	498

24% agreement, weighted kappa of 0.15.

CIOMS/RUCAM, Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method.

reappeared following unintentional re-exposure to the putative drug. Concordance between methods in assessing hepatotoxicity in cases with positive re-challenge was reached in 11% of ratings (three of 28). The correlation between scales in cases with poor outcome of DILI (death or liver transplantation) was also low (33% of ratings).

The differences in the percentage of agreement between-observers and between-scales are depicted in Figure 1 and from which it can be seen that the majority of the cases classified by the NADRPS scale as 'possible' by observer I were scored as 'probable' by observer II. Of 32 nondrug-related cases, only six ranked as 'unlikely' by the Naranjo scale, being the remainder classified as possible (Table 3). On the contrary, control cases were classified as 'unlikely' in the 98% of cases using the CIOMS scale. There was only a 15% concordance between scales in control cases (Table 4).

Cut-off scores of five points on the NADRPS scale and six points on the CIOMS/RUCAM scale were chosen to define a result as positive (drug related) or negative (not drug related). Based on these criteria, a 2 × 2 contingency table comparing results of both scales showed 54% sensitivity, 88% specificity, 95% of positive predictive value and a 29% negative predictive value for the NADRPS (Table 6).

DISCUSSION

Prompt recognition of culprit drugs as the cause of liver injury is the most important aspect in the management of hepatotoxicity.³⁰ To improve the evaluation of cases of suspected hepatotoxicity and provide

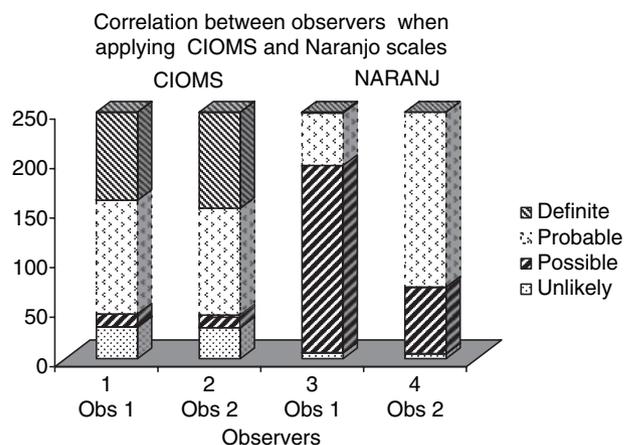


Figure 1. Comparison of categories of suspicion obtained by two observers applying the Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method and the Naranjo scales to 225 cases of drug-induced liver injury.

an objective and uniform approach for determining the likelihood of drug involvement, different diagnostic tools have been proposed. The qualities required for a good diagnostic tool are reproducibility and validity. Reproducibility of any scientific instrument is important when comparing results so that different users arrive at the same assessment despite differences in time and place. Validity implies that the method is able to distinguish between cases that are drug related and those that are not.

In this study, we compared a nonspecific method of causality assessment of adverse drug reactions (the NSDRPS scale) with a liver-specific, widely accepted standard (the CIOMS/RUCAM scale) in a large cohort of DILI cases reported in a national registry of hepatotoxicity. The low absolute agreement of 24% between scales (κ_w : 0.15) and the poor reliability and reproducibility of the NADRPS confirm previous experience from our group with respect to individual cases.^{13–15} This study also supports the advantage of CIOMS/RUCAM over the NADRPS method in the evaluation of cases of DILI.

The series of DILI cases evaluated in this study represent the last consecutive DILI cases included in the registry. A selection bias such as the inclusion of cases with a greater uncertainty of hepatotoxicity is, therefore, unlikely. Further, the demographic and clinical characteristics, type of liver injury and causative drugs in the group included in this study did not differ from those seen in the total cohort recorded in the Registry

		Naranjo ≥ 5 (definite/probable)	Naranjo < 5 (possible/unlikely)	Total
CIOMS/RUCAM ≥ 6	Definite/probable	220	189	409
CIOMS/RUCAM < 6	Possible/unlikely	11	78	89
Total		231	267	498

Table 6. 2×2 contingency table comparing mean Naranjo and CIOMS/RUCAM scores

Sensitivity: 0.54; specificity: 0.88; positive predictive value: 0.95; negative predictive value: 0.29.

CIOMS/RUCAM, Council for International Organizations of Medical Sciences/Roussel Uclaf Causality Assessment Method.

at the time of the analysis ($n = 461$).¹⁶ In addition, DILI cases used in this study were submitted to the Registry when specialist doctors had decided, on clinical criteria, that the hepatic reaction was linked to the culprit drug. They submitted the information in a structured report form, and some cases were also subjected to peer-review. These measures explain why so few cases assessed were excluded based on the CIOMS/RUCAM method. In addition, the observers who classified all cases using both clinical scales were two independent doctors with experience and motivation in hepatotoxicity and trained in the use of these scales before performing the study.

A detailed analysis of the different ratings shows that disagreements between observers in the application of the Naranjo scale were related to the use of clinical judgement in evaluating alternative aetiological explanations. These include nondrug and drug-related causes as a standardized methodology is not provided. This deficiency has been highlighted by the authors who developed the rating system.^{9, 31} Further, the scale is not organ specific, strict chronological criteria with respect to the type of liver injury are not defined, associated risk factors are not taken into account, and much is left to individual interpretation.^{15, 31} These factors may partly explain why performance of the NADRPS scale did not improve when type of liver damage or presence of hypersensitivity features was considered. It is worth pointing out that even definite criteria such as positive re-challenge showed disagreements among observers when answered. The NSDRPS scale has confusing questions (positive re-challenge and previous exposure/cross-reactivity), which are prone to different interpretations and responses. 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 or a positive re-challenge as definite evidence of hepatotoxicity. The Naranjo algorithm was designed for evaluation of adverse drug reactions related to pharmacological actions of the drugs; hence contains questions such as 'drug concentrations and monitoring', 'dose relationship', 'placebo response', which are clearly not relevant to idiosyncratic DILI. Therefore, we believe that there is no problem of consistency on observers' reading but with the scoring criteria that preclude a better categorical agreement from being obtained. Indeed, it would have been surprising, if this nonliver-specific scale turned out to be very useful in the causality assessment of hepatotoxicity, as Naranjo scale was designed for the evaluation of all types of adverse drug reactions. We consider that the weightings of criteria might differ among different adverse reactions to take into account the singularities of each therapeutic problem.

The principal advantages of the Naranjo ADR scale are simplicity, not time-consuming and wide applicability. It was initially validated using 63 published reports of adverse drug reactions, and was subsequently confirmed in 106 cases of nonliver-related events.^{9, 31} There was full agreement in 35% of cases ($r = 0.79$). The between-observer agreement in this study (45%; κ_w : 0.17) when evaluating hepatotoxicity was no better than the 50% agreement (κ_w : 0.30) found if two observers had evaluated general adverse reactions without using a diagnostic scale.³² As such, the Naranjo instrument does not add consistency or objectivity to the causality assessment of DILI cases assessed on clinical grounds.

This study shows a low sensitivity and low negative predictive value for the Naranjo scale and consequently a low capability of diagnosing DILI. Even cases with positive re-challenge (consensually accepted as the gold standard in the diagnosis of

DILI) or cases with liver biopsy results indicating 'compatible with DILI', are under-diagnosed by this scale. It is of note that only one case (positive rechallenge) of the 225 cases was scored as 'definite' on the Naranjo scale by one of the observers. These results highlight the difficulty of obtaining high scores when applying NADRPS criteria and underline the importance of providing the individual score for each item to assess the actual judgement of the reporter. Conversely, the ability of the NADRPS scale to exclude nondrug-related cases confidently was very limited, as very few cases were assigned this category. Hence, we could state that the Naranjo scale shows a low discriminative power and a limited capability to distinguish between adjacent categories, such as 'possible' and 'probable'.

A limitation of this study is the assumption that the CIOMS/RUCAM system can act as an external standard in evaluating the validity of the Naranjo scale. However, the CIOMS/RUCAM scale has been shown to perform well in evaluating DILI cases, not only in the present, but also in previous studies.¹² However, some modifications have been suggested to optimize the method. These include incorporating relevant risk factors when an appropriate setting is present (e.g. HIV infection in sulphonamide users, or valproic acid in younger individuals), or the inclusion of new items as drug-lymphocyte stimulation test³³ and eosinophilia. These would improve the evaluation of cases of immuno-

allergic DILI, while not excluding cases with a higher latency period.³⁴ Finally, future refinements in the weighting given to individual parameters based on statistical evaluations derived from extensive databases are to be performed.³⁵

In conclusion, we discourage the application of the Naranjo scale in the causality assessment of suspected drug-induced liver impairment. Although in need of further refinement, the CIOMS/RUCAM scale provides an optimal level of objectivity. It appears to be more reliable and reproducible and could be of considerable clinical value in assessing complex patients, and in research settings. Further, the scale can be 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.

ACKNOWLEDGEMENTS

Declaration of personal interests: None. *Declaration of funding interests:* This study was supported, in part, by research grants from the Agencia Española del Medicamento and Fondo de Investigación Sanitaria (FIS 04-1688, FIS 07-0980). The funding sources had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report and in the decision to submit the manuscript for publication.

REFERENCES

- Dossing M, Sunne J. Drug induced hepatic disorders. Incidence, management and avoidance. *Drug Saf* 1993; 9: 441-9.
- Ostapowicz G, Fontana RJ, Schiodt FV, et al. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann Intern Med* 2002; 137: 947-54.
- Davis M, Williams R. Hepatic disorders. In: Davies DM, ed. *Textbook of Adverse Drug Reactions*. Oxford: Oxford University Press, 1977: 146-72.
- Bakke OM, Manocchia M, De Abajo F, et al. Drug safety discontinuations in the United Kingdom, the United States, and Spain from 1974 through 1993: a regulatory perspective. *Clin Pharmacol Ther* 1995; 58: 108-17.
- Navarro VJ, Senior JR. Drug-related hepatotoxicity. *N Engl J Med* 2006; 354: 731-9.
- Sgro C, Clinard F, Ouazir K, et al. Incidence of drug-induced hepatic injuries: a French population-based study. *Hepatology* 2002; 36: 451-5.
- Aithal GP, Rawlins MD, Day CP. Accuracy of hepatic adverse drug reaction reporting in one English health region. *BMJ* 1999; 319: 1541.
- Pere JC, Begaud B, Haramburu F, et al. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; 40: 451-561.
- Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; 30: 239-45.
- Danan G, Bénichou C. 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* 1993; 46: 1323-30.
- Bénichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs: II. An original novel or validation of drug causality assessment methods: case reports with positive rechallenge. *J Clin Epidemiol* 1993; 46: 1331-6.
- Lucena MI, Camargo R, Andrade RJ, et al. Comparison of two clinical scales for causality assessment in hepatotoxicity. *Hepatology* 2001; 33: 123-30.
- Andrade RJ, Guilarte J, Salmerón FJ, et al. Benzylpenicillin-induced pro-

- longed cholestasis. *Ann Pharmacother* 2001; 35: 783–4.
- 14 García-Cortés M, Lucena MI, Andrade RJ, *et al.* Lansoprazole-induced hepatic dysfunction. *Ann Pharmacother* 2003; 37: 1731.
 - 15 Garcia-Cortes M, Lucena MI, Andrade RJ, *et al.* Is the Naranjo probability scale accurate enough to ascertain causality in drug-induced hepatotoxicity? *Ann Pharmacother* 2004; 38: 1540–1.
 - 16 Andrade RJ, Lucena MI, Fernandez MC, *et al.* Spanish Group for the Study of Drug-induced Liver Disease. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. *Gastroenterology* 2005; 129: 512–21.
 - 17 Gavilán JC, Bermúdez FJ, Salgado F, *et al.* Fitoterapia y hepatitis. *Rev Clin Esp* 1999; 199: 693–4.
 - 18 Gallardo-Caballero E, Lucena MI, Andrade RJ, *et al.* Severe toxic hepatitis related to chlormethiazole. *J Hepatol* 2003; 39: 1092–3.
 - 19 Rodrigo L, de Francisco R, Perez-Pariente JM, *et al.* Nimesulide-induced severe hemolytic anemia and acute liver failure leading to liver transplantation. *Scand J Gastroenterol* 2002; 37: 1341–3.
 - 20 Lopez-Torres E, Lucena MI, Andrade RJ, *et al.* Tetrabamate-induced hepatotoxicity. Report of seven cases and literature review. *Gastroenterol Hepatol* 2002; 25: 589–93.
 - 21 Lucena MI, Andrade RJ, Rodrigo L, *et al.* Trovafloxacin-induced acute hepatitis. *Clin Infect Dis* 2000; 30: 400–1.
 - 22 Lucena MI, Andrade RJ, Vicioso L, *et al.* Prolonged cholestasis after raloxifene and fenofibrate interaction: a case report. *World J Gastroenterol* 2006; 12: 5244–6.
 - 23 Verge C, Lucena MI, López-Torres E, *et al.* Adverse hepatic reactions associated with calcium carbimide and disulfiram therapy: is there still a role for these drugs? *World J Gastroenterol* 2006; 12: 5078–80.
 - 24 Lucena MI, Andrade RJ, Fernández MC, *et al.* Determinants of the clinical expression of amoxicillin-clavulanate hepatotoxicity: a prospective series from Spain. *Hepatology* 2006; 44: 850–6.
 - 25 Martindale W. The complete drug reference. 33rd edn. London: Pharmaceutical Press, 2002.
 - 26 Benichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. *J Hepatol* 1990; 11: 272–6.
 - 27 World Health Organization Collaborating Centre for Drugs Statistics Methodology, Anatomical Therapeutic Chemical (ATC) Classification Index Including Defined Daily Dose (DDDs) for Plain Substances. Oslo: World Health Organization Collaborating Centre for Drugs Statistics Methodology, 2002.
 - 28 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–20.
 - 29 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
 - 30 Zimmerman HJ. *Hepatotoxicity. The Adverse Effects of Drugs and other Chemicals on the Liver*, 2nd edn. Philadelphia: Lippincott Williams & Wilkins, 1999.
 - 31 Lanctôt KL, Naranjo CA. Comparison of the Bayesian approach and a simple algorithm for assessment of adverse drug events. *Clin Pharmacol Ther* 1995; 58: 692–8.
 - 32 Naranjo CA, Busto U, Abel JG, *et al.* Empiric delineation of the probability spectrum of adverse drug reactions. *Clin Pharmacol Ther* 1981; 29: 267–8.
 - 33 Kaplowitz N. Idiosyncratic drug hepatotoxicity. *Nat Rev Drug Discov* 2005; 4: 489–99.
 - 34 Takikawa H, Takamori Y, Kumagi T, *et al.* Assessment of 287 Japanese cases of drug induced liver injury by the diagnostic scale of the International Consensus Meeting. *Hepatol Res* 2003; 27: 192–5.
 - 35 Kaplowitz N. Causality assessment versus guilt-by-association in drug hepatotoxicity. *Hepatology* 2001; 33: 308–10.

APPENDIX

The Spanish Group for the Study of Drug-induced Liver Disease (GEHAM)

Participating clinical centres and their principal investigators: Hospital Universitario Virgen de la Victoria, Málaga (Coordinating Centre): R. J. Andrade, M. I. Lucena, R. Camargo, E. García-Ruiz, M. García-Cortés, R. González-Grande, M. R. Cabello, R. Alcántara, E. Lopez-Torres, K. Pachkoria, M. J. Puche and Y. Borraz; Hospital Torrecárdenas, Almería: M. C. Fernández, G. Peláez, M. Casado, J. L. Vega, F. Suárez, M. Torres and M. González-Sánchez; Hospital Virgen Macarena, Sevilla: J. A. Durán, M. Jiménez-Sáez and M. Villar; Hospital Universitario Virgen de Valme, Sevilla: M. Romero-Gomez and A. Madrazo; Hospital Central de Asturias, Oviedo: L. Rodrigo-Saez, V. Cadahía and R. De Francisco; Hospital de Puerto

Real, Cádiz: J. M. Pérez-Moreno and M. Puertas; Hospital Universitario San Cecilio, Granada: J. Salmerón and A. Gila; Hospital Germans Trias i Puyol, Barcelona: R. Planas, M. I. Barriocanal, N. López-Rodríguez, E. Montaner, F. García-Góngora and J. Costa; Hospital Universitario Virgen de las Nieves, Granada: R. Martín-Vivaldi and F. Nogueras; Hospital Costa del Sol, Málaga: J. M. Navarro and J. F. Rodríguez; Hospital La Inmaculada, Huércal-Overa, Almería: H. Sánchez-Martínez; Hospital Puerta del Mar, Cádiz: F. Díaz, M. J. Soria and C. Martínez-Sierra; Hospital Reina Sofía, Córdoba: J. L. Montero, E. Fraga and M. De la Mata; Hospital 12 de Octubre, Madrid: T. Muñoz-Yagüe and J. A. Solís-Herruzo; Hospital Marqués de Valdecilla, Santander: F. Pons and R. Taheri; H. Sant Pau, Barcelona: C. Guarner and D. Monfort; Hospital Carlos Haya, Málaga: M. Jiménez-Pérez; Hospital Xeral-Calde, Lugo: S. Avila-Nasi; Hospital de Donosti, San Sebastián: M. García-Bengoechea;

Hospital de Mendaro, Guipuzcuoa: A. Castiella; Hospital de Basurto, Vizcaya: S. Blanco; Hospital Clinic, Barcelona: M. Bruguera; Hospital Morales Messeguer: H. Hallaf; Hospital Clínico Universitario Miguel Servet, Zaragoza: M. A. Simón; Hospital Juan Ramón Jiménez, Huelva: M. Ramos and T. Ferrer; Hospital Ciudad de Jaén: E. Baeyens; Hospital de Osuna, Sevilla: J. Pérez-Martínez; Hospital General Básico de Vélez, Málaga: F. Santalla and C. Sánchez-Robles; Hospital Gregorio Marañón, Madrid: R. Bañares; Hospital General de Valencia: M. Diago; Hospital Sagunto, Valencia: J. Primo and J. R. Molés.