Analysis of causality assessment methods in suspected HILI cases shows relevant gaps as assessed for accuracy, bias and transparency using data provided through the Freedom of Information Act (FOIA)

Cyril M. Willson

EuSci LLC, Gretna, NE, USA

Correspondence to: Cyril M. Willson. EuSci LLC, 1309 S 204th St, #293, Elkhorn, NE 68022, USA. Email: cmwillson@gmail.com.

Abstract: Causality Assessment Methods (CAMs) are employed to assess adverse effects from various agents. The use of the United States based Drug Induced Liver Injury Network (DILIN) method, which relies on global introspection (GI) or expert opinion (EO) for causality assessment in the case of herb-induced or drug-induced liver injury (HILI/DILI) is examined for bias. The accuracy of the Roussel Uclaf Causality Assessment Method (RUCAM) as used by the DILIN, overall transparency of the DILIN's use of both CAMs and the ability of the RUCAM to resist bias are also assessed. Data obtained from a Freedom of Information Act (FOIA) production by the National Institutes of Health (NIH) for material related to a publication by the US DILIN were analyzed to determine if any of 10 chosen forms of cognitive dispositions to respond (CDRs) or cognitive bias were present in an investigation of a dietary supplement, OxyELITE Pro (OEP). Data not originally included for publication were also utilized to assess the accuracy of the DILIN's RUCAM scoring and transparency. To assess the RUCAM's possible resistance to bias, mean RUCAM scores were calculated to evaluate those produced by a Primary Investigator (PI) versus computer for OEP and non-OEP products. A minimum of 4 and up to 10 CDRs were present. The data also showed the RUCAM may resist bias as there was no difference in causality grading between the mean PI and computer-based RUCAM scores. However, the lack of inferential analyses and small sample size are limitations. RUCAM scores by DILIN authors for OEP consisted of 1 as "unlikely" 4 as "possible" and 2 as "probable." However, when scores were recalculated based upon previously unreported data, RUCAM scores decreased substantially with 3 that should have been "excluded" and 4 as "possible," indicating inaccurate scoring. Discrepancies between published data and those obtained via FOIA showed a lack of transparency. It is concluded that the DILIN method lacks transparency while being prone to bias. The RUCAM is the most appropriate method for evidence-based medicine but requires data to be reported objectively and transparently in order to avoid inaccurate scoring, misdiagnoses and incorrect causality attribution.

Keywords: Causality assessment; drug-induced liver injury; herb-induced liver injury; hepatotoxicity; bias; cognition

Introduction

In the realm of adverse effects from various agents (e.g., pharmaceuticals and dietary supplements) causality assessment methods (CAMs) may be used to assist clinicians and others in determining the likelihood that an agent caused an adverse effect. Ultimately however, because the data evaluated are case reports, rather than provide an objective measure of a cause-effect relationship, CAMs provide a way to measure the assessor's uncertainty or

degree of belief in the relationship (1,2). Herb-induced liver injury (HILI) or drug-induced liver injury (DILI) are rare but complex diseases that have been assessed for causality by the most commonly used Roussel Uclaf Causality Assessment Method (RUCAM) or alternatively by the US Drug Induced Liver Injury Network (DILIN) method (3). These methods differ substantially from each other. For instance, the RUCAM utilizes an objective algorithm that provides a final causality score and grading based on key components that are defined and scored individually (3,4). The RUCAM is also validated using an external reference standard (i.e., positive re-challenge cases) and presents its results in a transparent manner (3,4). Conversely, the DILIN method utilizes global introspection (GI), also known as expert opinion (EO) which has long been noted to suffer from serious limitations, including bias, subjectivity, inter and intra-observer variability and lack of transparency (2,5-12). The DILIN method also relies upon non-defined components which are not scored individually, while utilizing an obscure causality scoring which is graded by percentage and finally, it is not validated (3). Ruling out alternative causes transparently and systematically as it is done in the RUCAM is particularly important in the case of HILI/DILI, as it is a diagnosis of exclusion since other causes of acute and chronic liver disease may mimic the condition (12,13). More recently those within the network of the DILIN have claimed that the DILIN method is superior to the RUCAM (14). The EO-based process utilized by the DILIN method was intended to decrease inter-observer variability and minimize bias, which is said to be accomplished by the use of three EOs which must reach a consensus for scoring causality (15). However, such a method has not been assessed to examine what ability, if any, it has to reduce bias. EO is inherently subjective and is consistently considered the lowest form of evidence (16,17), thus the notion that having two additional experts would reduce bias to any significant extent is questionable, particularly when preconceived notions are pervasive and humans, including experts are poor at assessing a cause-effect relationship objectively and validly (2,18). Others have noted that EO itself is not a form of evidence but rather, a judgement of it (19). While EO is still a cornerstone in the diagnosis of most conditions, it is often aided or confirmed with a validated and objective biomarker, bioassay, bacteriological test or chemical assay. In the case of HILI/DILI, there is no such test currently available (11,12). A recent case series of HILI published by the DILIN became a matter of dispute due to potential confounders such as co-

medication, alternative liver diseases, incorrect causality attribution and inaccurate, misleading causality grading associated with a lack of transparent case data presentation and causality grading methodology (20,21), while the association for the product in question, OxyELITE Pro (OEP) with liver injury in general, has been controversial (20-31). The impression prevails that DILIN assessments are opaque, not allowing peers the opportunity to assess the validity of the work. However, information obtained from a Freedom of Information Act (FOIA) production by the National Institutes of Health (NIH) provides a window for insight into the DILIN methodology. The aims of the analyses presented herein are to assess the use of the RUCAM by the DILIN for accuracy and overall transparency in the use of both CAMs; assess the level of cognitive biases in the DILIN method; evaluate the ability of the RUCAM to resist bias, and critically evaluate claims of superiority of the DILIN method versus the RUCAM.

Methods

Data gathered from a FOIA production by the NIH were utilized to evaluate the use of the RUCAM by the DILIN for accuracy by comparing the published RUCAM scores of a DILIN publication (20) with RUCAM scores obtained after consideration of data not originally included for publication; transparency was assessed by comparing information included by the DILIN publication with information obtained via FOIA. RUCAM scores for potential alternative causes were also calculated. To assess potential bias in the DILIN method, clinical data and commentary were analyzed for cognitive dispositions to respond (CDRs) or cognitive bias; definitions for CDRs were taken from previous publications and applied as it relates to causality assessments (32,33) [see section *Cognitive dispositions to respond (cognitive bias) in cases 1–7* and *Definitions*]. CDRs are a form of cognitive error that can lead to diagnostic error. Out of 33 forms of CDRs defined in the literature, 10 were identified as being relevant to case data. To evaluate the RUCAM's ability to resist bias, mean RUCAM scores were calculated for those given by a primary investigator (PI) and those obtained via computer, to evaluate potential differences between a presumably more objective method of RUCAM scoring (i.e., via computer) versus a human PI. The PI was identified as the responsible party providing enrollment of the patient. Finally, a critical evaluation of the claimed superiority of the DILIN method versus the RUCAM was performed by evaluating published

Table 1 RUCAM scores of listed and omitted medications for cases 1–7

Case #	Medications listed	Medications omitted [RUCAM score by investigators]	Listed medications [RUCAM score by investigators]	All medications [omitted computerized RUCAM score]
1	OEP; Celsius; whey protein	Ibuprofen [ND]	OEP [2]; Celsius [5]; whey protein [ND]. OEP, Celsius and whey protein initially given scores of 7† by PI	Ibuprofen [4]; OEP [3]; whey protein [5]; Celsius [5]
2	Stacker 3, OEP	Alprazolam [ND]; zolpidem [ND]	Stacker 3 [4]†; OEP [5]	Stacker 3 [4]; OEP [4]; alprazolam [2]; zolpidem [4]
3	OEP	Levonorgestrel/ethinyl estradiol [ND]; multivitamin [ND]; norethisterone [ND]	OEP [7]	OEP [5]; multivitamin [3]; norethisterone [4]; levonorgestrel/ethinyl estradiol [ND]
4	None	NA	OEP [5]	NA
5	TMP/SMZ, OEP	Xylitol [ND], naproxen [ND], zolpidem [ND]; acetylsalicylic acid [ND]; celecoxib [ND], fluvoxamine [ND], unknown [ND]	TMP/SMZ [6]†; OEP [6]	TMP/SMZ [7]; OEP [4]; naproxen [0]; zolpidem [0]; acetylsalicylic acid [0]; celecoxib [0], fluvoxamine [1], unknown [0]
6	Azithromycin; OEP; turmeric; flaxseed oil	Alkaline booster [ND], Mucinex fast max cold and sinus [ND], bupropion [ND], ethinyl estradiol/norgestimate [ND], venlafaxine [ND]	Azithromycin [ND]; OEP [4]; OEP and azithromycin originally scored at RUCAM of 2† by PI; Turmeric [ND]; Flaxseed Oil [ND]	ND
7	OEP; ravage; hydroxycut; N.O. Xplode; creatine	Ibuprofen [ND]; whole body vitamin pack [1]	Ravage [1]†; OEP [3]; OEP originally scored RUCAM of 1† by PI; Hydroxycut [1]†; N.O. Xplode [ND]; creatine [1]†	Ravage [4]; OEP [4]; Hydroxycut [4]; ibuprofen [4]; whole body vitamin pack [4]; creatine [4]

RUCAM total score with resulting causality grading: ≤0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ≥9, highly probable. †, scores not reported in original publication by the DILIN (20). NA, Not Available; ND, not done; OEP, OxyELITE Pro; PI, primary investigator; RUCAM, Roussel Uclaf Causality Assessment Method; TMP/SMZ, trimethoprim/sulfamethoxazole.

literature and the analyses included herein.

Results

Overall, the data from a FOIA production demonstrated bias for the DILIN method, inaccurate RUCAM scoring caused by incorrect accounting of time to onset, course of the reaction, concomitant agents, viral exclusion and re-challenge, with scoring generally shifted upwards towards OEP and downwards for alternative causes. Additionally, RUCAM assessments for alternative agents were often either not performed or not reported. *Tables 1,2,3,4* show discrepancies between data obtained via FOIA and those published by the DILIN indicating a lack of transparency overall. Regarding the DILIN method, several instances of CDRs or cognitive bias for each of the 7 published cases where OEP was implicated were identified (*Table 5*), with a minimum of 4 and up to all 10 of the chosen forms of cognitive bias present. A

comparison of mean RUCAM scores generated by the expert PI versus a computer for OEP and non-OEP product was performed (*Table 6*), which showed that while the RUCAM scores for OEP by the PI were higher than computer-based scores, non-OEP RUCAM values were scored lower by the PI versus the computer indicating a potential bias towards implicating OEP compared to alternative causes; however, the mean scores for a PI versus computer did not result in a change of causality grading indicating the RUCAM may resist bias. With respect to accuracy, a comparison between RUCAM scores for OEP originally published by the DILIN and recalculated scores based upon unpresented data is made (*Tables* 7 and *8*, respectively) and shows a substantial decrease in scores after consideration of those data, reflecting a general lack of accuracy and transparency. Finally, RUCAM scoring for alternatively implicated agents is also given (*Table 9*), which shows agents with equal or significantly greater RUCAM scores.

Table 2 Liver injury onset dates and medication start and stop dates for cases 1–7

Case #	Liver injury onset date	Medication start date of use	Medication stop date of use	Comments
1	08/16/2011	Ibuprofen—08/9/2011	Ibuprofen—08/12/2011	Celsius contained Green Tea Extract. Peak ALT of 8,384 U/L on 08/16/2011, >50% decrease in difference between peak ALT and ULN to 3,428 U/L on 8/18/2011
		Celsius—06/15/2011	Celsius—08/11/2011	
		OEP—05/8/2011	OEP—06/13/2011	
		Whey Protein—07/18/2011	Whey Protein—08/11/2011	
2	05/07/2013	Stacker 3—2010	Stacker 3—04/2013	Stacker 3 had various formulation changes from 2010 through 2013. PI noted current version contained Green Tea Extract. Peak ALT of 1,583 U/L on 05/07/2013, >50% decrease in difference between peak ALT and ULN to 754 U/L on 05/25/2013
		OEP—01/2011	OEP—03/2011	
		OEP—01/2013†	OEP—04/2013†	
		Alprazolam—2011	Alprazolam—continued	
		Zolpidem—2011	Zolpidem—04/2013	
3	06/15/2013	Levonorgestrel/ethinyl estradiol—05/2013	Levonorgestrel/ethinyl estradiol—06/15/2013	Patient had previously used OEP for nearly 2 years prior to cessation due to anticipated pregnancy with no adverse effect. Patient resumed taking OEP after birth. Peak ALT of 2,242 U/L on 06/20/2013, >50% decrease in difference between peak ALT and ULN to 754 U/L on 06/28/2013
		Multivitamin—05/2012	Multivitamin—05/2013	
		Norethisterone—12/29/2012	Norethisterone—06/15/2013	
		OEP—07/2010	OEP—04/2012	
		OEP—04/2013	OEP—06/14/2013	
4	08/2013	OEP—07 or 08/2011	OEP—08/2013	OEP was consumed >2 years without issue
5	10/24/2013	OEP—11/4/2011	OEP—10/14/2013	OEP consumed intermittently >2 years without adverse effects. Patient claims eyes occasionally became icteric. Patient was poor historian of herbal use or when and how often she took OEP. Did not believe she took Super Thermogenic Formula. Naproxen and Acetylsalicylic acid were taken as needed. Self-reported use of TMP/SMZ 2-3 times prior without issue. Peak ALT of 1772 U/L on 10/24/2013, >50% decrease in difference between peak ALT and ULN to 824 U/L on 10/28/2013
		TMP/SMZ—10/15/2013	TMP/SMZ—10/22/2013	
		Xylitol—06/2013	Xylitol—10/24/2013	
		Naproxen—NA	Naproxen—NA	
		Zolpidem—2003	Zolpidem—continued	
		Celecoxib—2003	Celecoxib—continued	
		Acetylsalicylic acid—NA	Acetylsalicylic acid—NA	
		Fluvoxamine—09/12/08	Fluvoxamine—continued	
		Unknown—NA	Unknown—NA	
6	11/12/2013	Alkaline booster—06/15/13	Alkaline booster—10/30/13	Liver Tests were previously normal but were noted to be mildly elevated at a primary care visit on 10/29/13 and continued to worsen on 11/08/13. Prior use of OEP for >120 days and onset occurring only after azithromycin course is suggestive of the latter as a cause. Peak ALT of 729 U/L on 11/19/2013, >50% decrease in difference between peak ALT and ULN to 88 U/L on 12/03/2013
		Mucinex Fast Max Cold and Sinus—10/15/13	Mucinex Fast Max Cold and Sinus—10/19/13	
		Bupropion—2003	Bupropion—continued	
		Ethinyl estradiol/ norgestimate—2010	Ethinyl estradiol/ norgestimate—continued	
		Venlafaxine—2005	Venlafaxine—continued	
		Turmeric—01/1/13	Turmeric—10/29/13	
		Flaxseed oil—06/ 2012	Flaxseed oil—10/29/13	
		OEP—6/15/13	OEP—10/30/13	
		Azithromycin—10/14/13	Azithromycin—10/24/13	

Table 2 (*continued*)

Table 2 (*continued*)

Case #	Liver injury onset date	Medication start date of use	Medication stop date of use	Comments
7	12/03/2013	OEP—03/15/13	OEP—10/15/13	Ravage and Whole Body Vitamin Pack contain two known hepatotoxic agents. Peak ALT of 3,404 U/L on 12/03/2013, >50% decrease in difference between peak ALT and ULN to 971 U/L on 12/08/2013
		Ravage—04/15/13	Ravage—09/15/13	
		Hydroxycut—05/15/12	Hydroxycut—05/15/13	
		Whole Body Vitamin Pack—04/15/13	Whole Body Vitamin Pack—11/15/13	
		Creatine—05/15/11	Creatine—04/15/13	
		Ibuprofen—12/01/13	Ibuprofen—12/03/13	
		N.O. Xplode—NA	N.O. Xplode—NA	
		Entecavir—unclear	Entecavir—unclear	

†, despite not possessing actual dates of use, an unsupported estimate of 126 days of use prior to DILI onset is given and cessation of use 7 days prior to DILI onset is given by the PI. ALT, alanine aminotransferase; NA, not available; OEP, OxyELITE Pro; PI, primary investigator; TMP/SMZ, trimethoprim/sulfamethoxazole; U/L, units per liter; ULN, upper limit of normal.

Table 3 Summary of OxyELITE pro (OEP) use and reported and actual preexisting conditions for cases 1–7

Case #	Version of OEP consumed confirmed?	Version of OEP used	Preexisting conditions reported	Actual preexisting conditions
1	Yes	Old formula with DMAA	None	None
2	No	Likely old formula with DMAA	Manuscript states "no history of alcohol abuse"	Fatty liver via CT scan. Heterogeneous hepatic echotexture. History of alcohol abuse until 2011. History of alcohol abuse, PTSD, anxiety and eating disorder noted only in supplementary material
3	No	Likely old formula with DMAA	Obesity, mood disorder, prior gestational diabetes	Obesity, mood disorder, prior gestational diabetes
4	No	Likely old formula with DMAA	Hepatic steatosis	Hepatic steatosis
5	No	Likely old formula with DMAA	None	Depression, degenerative joint disease
6	No	Unknown	History of Depression	Sinusitis, dysthymia, headache, gastritis
7	No	Likely old formula with DMAA	History of "moderate alcohol use," chronic HBV Infection	Migraine/headache, dysthymia, alcohol-related liver disease (January, 2012)

CT, computed tomography; DMAA, 1,3-dimethylamylamine; HBV, hepatitis B virus; PTSD, post-traumatic stress disorder.

Analysis of HILI cases and causality assessment methods

Case 1

In the published work of the DILIN, this case was uncharacteristically included among the other 6 cases. The authors hypothesize that a change in the formulation of OEP in March of 2013, ultimately led to liver injury. However, Case 1 experienced liver injury in August of 2011, 2 years before the formula change occurred and thus is inconsistent with the authors' hypothesis. Furthermore, the patient was also consuming a green tea extract-containing product named "Celsius," which was considered the more likely culprit with DILIN and RUCAM scores of "probable"

Table 4 Alternative causes reportedly and actually excluded for cases 1–7

Case #	Alternative causes excluded as reported	Alternative causes actually excluded	Alternative causes not excluded	Comments
1	AIH, HAV, HBV, HCV, HEV	AIH, CMV, EBV, HCV	HAV†, HBV, HCV‡, HEV, HSV, VZV	
2	AIH, Ethanol, HAV, HBV, HCV, HEV	CMV, HAV, HBV, HCV, HEV	AIH§, EBV, Ethanol, HSV, VZV	
3	AIH, HAV, HBV, HCV, HEV	AIH, CMV, EBV, HAV, HBV, HCV, HEV	HSV, VZV	
4	AIH, HAV, HBV, HCV, HEV	HBV, HCV	AIH, CMV, EBV HAV, HEV, HSV, VZV	
5	AIH, HAV, HBV, HCV, HEV	AIH, CMV, HAV, HBV, HCV, HEV	EBV, HEV¶, HSV, VZV	
6	AIH, HAV, HBV, HCV, HEV	AIH, HBV, HCV	CMV, EBV, HAV†, HEV, HSV, VZV	Anti-HAV total was positive
7	AIH, HAV, HCV, HEV	AIH, HAV	CMV, EBV, Ethanol, HCV‡, HEV¶, HSV, VZV	HCV RNA not assayed until after transplant. Was negative post-transplant, but positive 6 months later

†, anti-HAV IgM not assayed. ‡, HCV RNA not assayed. §, the patient had several characteristics seen in AIH, including female sex, positive for antinuclear antibodies at 1:320, rising to 1:1,280 and an ALP: AST/ALT ratio less than 1.5; IgG and AMA were not assayed. ¶, HEV IgM not assayed. AIH, autoimmune hepatitis; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HAV, hepatitis A virus; HBV, hepatitis B virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HSV, herpes simplex virus; RNA, ribonucleic acid; VZV, varicella zoster virus.

Table 5 CDR determined for each case alongside DILIN scores for agents consumed by patients

Case #	CDR or cognitive bias that may lead to diagnostic error	DILIN score
1	CB, IC, MAB, UP	Ibuprofen [ND], Celsius [3], OEP [4], whey protein [5]
2	AB, CB, IC, MAB, OB, UP	Stacker 3 [4], OEP [2], alprazolam [ND], zolpidem [ND]
3	AB, CB, DM, IC, MAB, OB, PC, UP	Levonorgestrel/ethinyl estradiol [ND], multivitamin [ND], norethisterone [ND], OEP [1]
4	AB, CB, DM, IC, MAB, OB, PC, UP	OEP [2]
5	AB, CB, DM, IC, MAB, OB, PPE, PC, RR, UP	Xylitol [ND], naproxen [ND], zolpidem [ND], acetylsalicylic acid [ND], celecoxib [ND], fluvoxamine [ND], unknown [ND], OEP [3], TMP/SMZ [4]
6	AB, CB, DM, IC, MAB, OB, PC, UP	Alkaline booster [ND], mucinex fast max cold and sinus [ND], bupropion [ND], ethinyl estradiol/norgestimate [ND], venlafaxine [ND], turmeric [ND], flaxseed oil [ND], OEP [3], azithromycin [4]
7	AB, CB, DM, IC, MAB, OB, PC, UP	OEP [3], ravage [4], hydroxycut [5], whole body vitamin pack [ND], creatine [ND], ibuprofen [ND], N.O. Xplode [ND]

DILIN expert opinion causality scoring: 1 (definite: ≥95% likelihood), 2 (highly likely: 75–94% likelihood), 3 (probable: 50–74% likelihood), 4 (possible: 25–49% likelihood) to 5 (unlikely: <25% likelihood). AB, availability bias; CB, confirmation bias; CDR, cognitive dispositions to respond; DM, diagnosis momentum; DILIN, Drug Induced Liver Injury Network; IC, illusory correlation; MAB, multiple alternatives bias; ND, not done; OB, overconfidence bias; OEP, OxyELITE Pro; PPE, posterior probability error; PC, premature closure; RR, representativeness restraint; TMP/SMZ, trimethoprim/sulfamethoxazole; UP, unpacking principle.

and "possible," respectively (*Tables 1,2,5*). The patient had also been using high doses of ibuprofen just prior to liver injury onset but after symptoms arose. The PI stated that this case should not be included in the case series, yet it was included and assigned a DILIN causality score of "possible" while the PI-derived RUCAM score indicated it was "unlikely." Specifically, the onset of hepatitis (i.e., as measured via rise in alanine aminotransferase—ALT)

Table 6 Mean and standard deviation for OEP and Non-OEP RUCAM scores as determined by PI and computer

	OEP PI RUCAM	OEP computer RUCAM	Non-OEP PI RUCAM	Non-OEP computer RUCAM
Mean	4.6 †	4†	4‡	5‡
SD	2.07†	0.71†	2.16‡	1.41‡

RUCAM total score with resulting causality grading: ≤0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ≥9, highly probable. †, only scores for Cases 1–3, 5 and 7 were available for both parameters. ‡, only scores for Cases 1, 2, 5 and 7 were available for both parameters. OEP, OxyELITE Pro; PI, Primary Investigator; RUCAM, Roussel Uclaf Causality Assessment Method; SD, standard deviation.

Table 7 OEP scoring based on original RUCAM as scored by the DILIN (20)

Case	Time to onset	Course	Risk factors	Concomitant drugs	Search for non-drug causes	Previous information on hepatotoxicity	Response to re-administration	Total
1	1	0	0	−1	1	1	0	2
2	2	2	0	−1	1	1	0	5
3	2	2	0	0	2	1	0	7
4	1	2	0	0	1	1	0	5
5	1	3	1	−2	1	1	1	6
6	1	2	0	−1	1	1	0	4
7	1	0	1	−1	1	1	0	3

RUCAM total score with resulting causality grading: ≤0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ≥9, highly probable. OEP, OxyELITE Pro; RUCAM, Roussel Uclaf Causality Assessment Method.

Table 8 OEP rescored based on unreported information

Case	Time to onset	Course	Risk factors	Concomitant drugs	Search for non-drug causes	Previous information on hepatotoxicity	Response to re-administration	Total
1	Incompatible							Unrelated
2	NA or 1†	2‡	0	-2	1	1	0	3†‡
3	2¶	2	0	-2	2	1	0	5
4	1	2	0	0	0	1	0	4
5	1	2	1	-2	1	1	0	4
6	1	0§	0	-2	0	1	0	0§
7	Incompatible							Unrelated

RUCAM total score with resulting causality grading: ≤0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ≥9, highly probable. †, exact dates of use are unknown, however, an unsupported estimate of 126 days is also given prior to DILI onset and cessation of use 7 days prior to DILI onset, thus the score could be either not assessable or 1. ‡, exact dates of use are unknown however an unsupported estimate of cessation of use 7 days prior to DILI onset is given by the PI. If correct, this would result in a score of 2, resulting in a total RUCAM score of 3. §, bloodwork was taken 11/19/2013 (peak ALT) and not again until 12/03/2013, thus it is unknown if a >50% decreased in the difference between peak ALT and ULN occurred within 30 days (i.e., bloodwork from 11/29/13 or prior). If it had, total RUCAM score would be 2. Current score is based upon available data. ¶, starting date is only given as 4/2013; however, even calculating a range of 4/01/13 to 4/30/13 to onset date of 6/15/13 still yields a range of 46–75 days, within the suggestive range of 5–90 days. NA, not assessable; OEP, OxyELITE Pro; RUCAM, Roussel Uclaf Causality Assessment Method.

Table 9 Alternative causes scored based on unreported information

Case	Time to onset	Course	Risk factors	Concomitant drug(s)	Search for non-drug causes	Previous information on hepatotoxicity	Response to re-administration	Total
1 (CS)	2	3	0	–1	0	1	0	5
2 (S3)	NA or 1†	2¶	0	–2§	1	1	0	3†¶
3 (OC-LEE)	2¶	2	0	–2§	2	2	0	6
4	–	–	–	–	–	–	–	–
5 (TMP/SMZ)	2	3	1	–2§	1	2	0	7
6 (AZN)	2	0¶	0	–2§	0	2	0	2¶
7 (WBVP)	1‡	0	1	–1	–3	1	0	–1

RUCAM total score with resulting causality grading: ≤0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ≥9, highly probable. †, exact dates of use are unknown, however, use of >90 days is also indicated prior to DILI onset and unsupported cessation of use 7 days prior to DILI onset, thus the score could be either not assessable or 1. ‡, WBVP falls 3 days outside of the 15 days required for the time to onset and should be excluded unless it is slowly metabolized. Due to unknown pharmacokinetic properties of the herbs and the closest agent in proximity to injury, the score is included here for thoroughness. §, OEP is scored as a concomitant drug for the sake of accepting DILIN's assertions. ¶, exact dates of use are unknown for S3, however an unsupported estimate of cessation of use 7 days prior to DILI onset is given. If correct, this would result in a score of 2, yielding a total RUCAM score of 3. For AZN, bloodwork was taken 11/19/2013 (peak ALT) and not again until 12/03/2013, thus it is unknown if a >50% decreased in the difference between peak ALT and ULN occurred within 30 days (i.e., bloodwork from 11/23/13 or prior). If it had, total RUCAM score would be 4. Current score is based upon available data. For OC-LEE, starting date is only given as 5/2013; however, even calculating a range of 5/01/13 to 5/31/13 to onset date of 6/15/13 still yields a range of 15–45 days, within the suggestive range of 5–90 days. ALT, alanine aminotransferase; AZ, azithromycin; CS, Celsius supplement; NA, not assessable; OC, oral contraceptive—levonorgestrel/ethinyl estradiol; OEP, OxyELITE Pro; RUCAM, Roussel Uclaf Causality Assessment Method; S3; Stacker 3 Supplement; TMP/SMZ, trimethoprim/sulfamethoxazole; ULN, upper limit of normal; WBVP, whole body vitamin pack.

occurred more than 2 months after cessation of OEP use, implicating either ibuprofen or the "Celsius" supplement and indicating that OEP was incompatible with liver injury and should not have received a RUCAM score but should have been assessed as simply, "unrelated" rather than a score of "1" for "time to onset" (34). Even ignoring this error, there was incomplete viral exclusion, leading to an incorrect, "search for non-drug causes" score for the OEP RUCAM (*Tables 4,7*) resulting in an incorrect total score of "2" rather than "1."

Case 2

Aside from the previous history of alcoholism (i.e., the main manuscript erroneously claims no patient had a history of alcohol abuse, while the supplementary file states a history of alcohol abuse for case 2), the authors also neglect to note that this patient had possible autoimmune hepatitis (AIH) and image-based fatty liver (*Tables 2,3,4*). The authors indicate the dietary supplement, Stacker 3 XPLC Extreme (Stacker 3), which contained green tea extract, was also consumed concomitantly with OEP. No medications are

listed; however, case 2 was consuming a benzodiazepine as well as zolpidem, which were assigned a computerized RUCAM score of "2" and "4," respectively. OEP and Stacker 3 were both assigned equivalent computerized RUCAM scores of "4" (possible), while Stacker 3 received an expert-generated RUCAM score of "4" (possible), yet these also are not presented in the manuscript (*Table 1*). The expert-generated RUCAM for OEP was "5" (possible), while the DILIN score was "highly likely." (*Tables 1,5*) However, the lack of consideration for green tea extract, a known hepatotoxic agent (34,35), caused an incorrect "concomitant drugs" score of "–1," as opposed to "–2" for the OEP RUCAM (*Tables 7,8*). Additionally, because the actual dates of use were unknown (*Table 2*), the "time to onset" for the RUCAM should have been recorded as "not assessable" rather than a score of "2," while the "course" should have only received a score of "0" as there were no data available for dates of use to determine if a >50% decrease in the difference between peak ALT and Upper Limit of Normal (ULN) values occurred within 30 days after cessation of use, all resulting in an incorrect total RUCAM score of "5" (*Tables 7,8*). Nonetheless,

unsupported estimates as entered by the PI for length of use (126 days) and cessation prior to DILI onset (7 days) are also scored, leading to a corrected total RUCAM score of "3" (*Table 8*).

Case 3

In this case, the oral contraceptive, which was initiated the prior month and continued until DILI onset, despite being well-known to cause varying forms of DILI, was completely omitted from the manuscript and ignored by the experts leading to an incorrect "concomitant drugs" score of "0" as opposed to "–2" for the OEP RUCAM resulting in an incorrect total score of "7" (probable) rather than "5" (possible) (34,36,37) (*Tables 1,2,7,8*). Indeed in case 3 and case 5, some of the most frequently implicated agents were consumed by the patients (25). Additionally, it is evident that case 3 was not consuming the version implicated in causing liver injury (*Table 2*). The DILIN score was the highest available as "definite" (*Table 5*).

Case 4

In this case, the extremely long latency is questionable. For example, it has been stated that DILI is uncommon when latency exceeds 6 months, yet cases 4 and 5 had latency more than twice this amount of time, while case 7 was also beyond this time (14). Case 4 also demonstrates the lack of sufficient viral exclusion where not only hepatitis A, but less commonly encountered viral causes were not excluded (34,38,39), leading to an incorrect, "search for non-drug causes" score for the OEP RUCAM of "1" as opposed to "0," resulting in an incorrect total score of "5" (possible) as opposed to "4" (*Tables 1,4,7,8*). The DILIN score was "highly likely" (*Table 5*).

Case 5

This patient indicated that she did not believe she used the reformulated version. Additionally, the expert-derived RUCAM score for OEP was elevated compared to the computer-based RUCAM, while the expert-derived RUCAM for trimethoprim/sulfamethoxazole (TMP/SMZ) was lower as compared to the computer-derived score (*Table 1*). The expert-derived RUCAM for TMP/SMZ was equivalent to the OEP score but only the latter was reported. This case also had incorrect scoring for "response to re-administration," or re-challenge; the authors indicated

a score of "1" but this was not based upon objective liver injury markers (e.g., ALT) as required by the RUCAM (34); instead, the score was based upon a story by the patient of taking OEP for over 2 years and claiming her eyes would occasionally become icteric (*Tables 2,7,8*). The "course" was also incorrectly scored as "3" as peak ALT values were obtained on 10/24/2013 and >50% decrease in the difference between peak ALT values and the ULN did not occur until 10/28/2013, which was >8 days after cessation (i.e., 10/14/2013), producing a score of only "2." These errors led to an incorrect total RUCAM score of "6" (probable) as opposed to "4" (possible) while the DILIN score was "probable" (*Tables 1,2,5,7,8*).

Case 6

In this case the PI had originally assigned OEP and azithromycin equivalent DILIN scores of 4 (probable) along with a RUCAM score of "2" (unlikely) but would later indicate that "the laboratory findings appear to be most compatible with DILI from OxyELITE Pro" (*Tables 1,5,7*). The other implicated agent, azithromycin did not receive a RUCAM score and was not properly considered in the "concomitant drugs" scoring for the OEP RUCAM, being scored only as a concomitant drug with compatible or suggestive time to onset ("-1") when it is a known hepatotoxic agent and should have received a score of "–2" (*Tables 7,8*). In addition, various viral causes including HAV, were not excluded leading to an incorrect "search for non-drug causes" score of "1" as opposed to "0" (*Tables 4,7,8*), while an acetaminophen-containing product (i.e., Mucinex Fast Max Cold and Sinus) was never considered or reported (*Table 1*). These issues led to an unsubstantiated upgrade of the original PI-calculated total RUCAM score from "2" (unlikely) (*Tables 1,7,8*) to an incorrect score of "4" (possible) (34). Additionally, since the "course" should have received a score of "0" as opposed to an incorrect "2" (i.e., considering the >50% decrease between the difference in peak ALT values and the ULN occurred beyond 30 days after cessation of use), the total RUCAM should have been "0" or "excluded."

Case 7

In this case, OEP was implicated (i.e., OEP DILIN score "probable"; RUCAM score "possible") over a supplement named, whole body vitamin pack (WBVP) containing known hepatotoxic ingredients despite the fact that

OEP had not been used for 49 days prior to DILI onset, while the WBVP (DILIN score not done, RUCAM score "unlikely") was used only 18 days prior to DILI onset, making it a much more likely culprit (*Tables 1,2,5*) and indicating that OEP should have been assessed as "unrelated" via the RUCAM (34). The PI originally assigned a RUCAM score of "1" (unlikely) to OEP, the same as WBVP but this was upgraded to a score of "3" (possible) without explanation or substantiation (*Tables 1,7*). Additionally, other supplements, also containing known hepatotoxic ingredients (e.g., Ravage) or formulas previously implicated in causing liver injury (e.g., NO Xplode) were minimized or not considered by the authors (*Tables 1,2*); the PI expert-derived RUCAM was also much lower for Ravage as compared to the computer-derived score. Unlike the ingredients in OEP, Ravage contains two known hepatotoxic agents (i.e., Chinese Skullcap Root Extract—*Scutellaria baicalensis* and *Acacia catechu*) (23,40-44). The WBVP is not scored or mentioned by the authors, yet it also contained Chinese Skullcap Root Extract (*Scutellaria baicalensis*) and *Acacia catechu*. Furthermore, the WBVP had a closer temporal association than OEP (*Table 2*). NO Xplode is a dietary supplement that has been previously implicated in causing liver injury, including by the DILIN (23,45). While NO Xplode is mentioned as being consumed in the supplementary material, it is not assessed for causality (*Tables 1,2*). It also appears this patient was using the old version of OEP rather than the reformulated version (46). Additionally, the patient was known to have a history of alcohol abuse, being diagnosed with alcohol-related liver disease in January of 2012, yet no steps were taken to exclude what if any role, alcohol may have played. To the contrary, the authors indicate that the patient only had a history of "moderate alcohol use" (*Table 3*). In addition to non-viral causes, case 7 demonstrates the strong possibility of an acute exacerbation of chronic hepatitis B along with possible superimposition with hepatitis C, which was not adequately excluded prior to transplantation (*Tables 3,4*). While the authors note the negative Immunoglobulin M antibody to hepatitis B core antigen (IgM anti-HBc), the positive hepatitis B surface antigen (HbsAg) and confirmed hepatitis B virus DNA (HBV DNA) of 32,000 IU/mL indicate this patient fit the criteria of an acute exacerbation of chronic hepatitis B infection as noted by one DILIN investigator and several publications (47-50). Certainly the HBV DNA level of 32,000 IU/mL (179,200 copies/mL) is well above the 20,000 IU/mL threshold for treatment and is considered "high" and an indication of clinically

meaningful, active replication, rather than the claim by the authors of this being a "low" level (51,52). Treatment with entecavir also supports this (*Table 2*). This patient also fit with a demographic which experiences acute exacerbation of chronic hepatitis more frequently (i.e., homosexual males) (48). Even ignoring the failure to score as "unrelated" ("time to onset" incorrectly scored as "1") the "search for non-drug causes" score of "1" was also incorrect (i.e., HBV is not excluded) considering the active HBV infection and exacerbation (*Tables 6-8*).

Global introspection or expert opinion versus RUCAM

While it is difficult to make quantitative comparisons between the RUCAM and the EO-based DILIN method due to differences in scoring systems, one may notice the DILIN method was more likely to implicate OEP and less likely to implicate alternative medications/herbs/supplements compared to the RUCAM (*Tables 1,5*). For example, the computerized RUCAM used by the authors, which would presumably be the most objective, rated the likelihood of causality for OEP only as "possible" in 5/7 cases (i.e., cases 1–3 and 7; cases 4 and 6 were not assessed via computerized RUCAM but were also indicated as "possible" by the PI). However, the DILIN method indicated that OEP was "probable" or greater in 6/7 cases (i.e., cases 2–7). This is significant as scores of "probable" and "possible" have been used to differentiate between "positive" and "negative" cases, respectively, while others have indicated that only "probable" or "highly probable" cases should be considered (4,15). Furthermore, in some cases the DILIN method resulted in a reduced likelihood of causality for competing causes. For example, in case 5, TMP/SMZ was rated only as "possible" by the DILIN method, while the computerized and PI-derived RUCAM were both "probable." Another example can be seen in the case of azithromycin use in case 6. The DILIN has previously studied an 18-person cohort of azithromycin-induced liver injury where all patients were rated with a likelihood of "probable" or higher and no patient was rated as "possible." However, despite case 6 having an essentially prototypical display of azithromycin-induced liver injury when compared to the cohort (53), the DILIN method rated the drug only as "possible." Other anomalies and inconsistencies with the scoring performed on agents in that cohort (e.g., TMP/SMZ, oral contraceptives, dietary supplements) which were also scored in the paper examined

here have been discussed previously (23). These data also confirm what seems to be a more frequent occurrence; published case studies for HILI/DILI often have neglected alternative causes or incorrectly attributed causality in a substantial number of cases, often nearing 50% (23,54). Finally, the current data show that while the RUCAM scores obtained by the PI were generally higher for OEP and lower for non-OEP products/alternative causes as compared to the computerized RUCAM scores (*Table 1*), the mean differences did not result in a change in causality grading (*Table 6*), indicating that the RUCAM may be able to withstand some level of bias and subjectivity on the part of an investigator as compared to the EO-based DILIN method. This also confirms what others have noted regarding the general agreement between human and computerized RUCAM scoring (4). However, it should be noted that the small sample size and lack of inferential analyses is a limitation of these data.

Cognitive dispositions to respond (cognitive bias) in cases 1–7

When evaluating the data for the seven cases, various cognitive biases (see *Definitions*) were noted which may have been the cause of the elevated causality scores for OEP and the lower scores for alternative agents, leading to incorrect causality attribution and misdiagnosis.

Definitions

- ❖ Availability bias (AB): a form of bias in which agents that readily come to mind or those recently encountered by the observer are determined to be a more likely or frequent cause.
- ❖ Confirmation bias (CB): a form of bias in which the observer tends to seek out evidence to confirm causality, while ignoring or minimizing evidence which disconfirms it.
- ❖ Diagnosis momentum (DM): a form of bias in which causality is eventually assumed by the observer based upon labels placed upon the agent by outside observers (e.g., patients, colleagues, media, or public announcements).
- ❖ Heuristics: mental shortcuts used to quickly make a decision or judgement.
- ❖ Illusory correlation (IC): a form of bias in which causality is assumed merely due to the temporal relationship between an agent and a disease without

consideration of disconfirming evidence.
- ❖ Multiple alternatives bias (MAB): a form of bias in which the observer, when faced with multiple possibilities for causality will simply analyze a smaller subset, resulting in inadequate consideration of other possibilities.
- ❖ Overconfidence bias (OB): a form of bias in which the observer believes they know more than they truly do and relies more upon opinion as compared to evidence.
- ❖ Posterior probability error (PPE): a form of bias in which the observer overemphasizes prior experience with a particular patient when determining the likelihood that a given agent played a causal role.
- ❖ Premature closure (PC): a form of bias in which the observer determines that an agent played a causal role before supportive evidence has been verified.
- ❖ Representative restraint (RR): a form of bias in which only prototypical manifestations of a disease are considered to include or exclude a given agent.
- ❖ Unpacking principle (UP): a form of bias in which the observer fails to gather all relevant information or potentially relevant information via limited history-taking or limited history-giving on behalf of the observer and patient, respectively.

AB

Case 2 demonstrated AB as OEP is referenced as a "known hepatotoxic product" and is assigned a higher causality score than Stacker 3, despite the fact that the latter contained a known hepatotoxic substance (i.e., green tea extract and possibly others), while OEP was used for 2 months previously without adverse effect (*Table 2*). AB is evident for cases 3–7 due to the reference of OEP being, at the time, recently implicated in causing liver injury as a justification for substantially elevating its potential causal role.

CB

Case 1 demonstrated CB as OEP, despite lacking any temporal connection to the liver injury and consisting of an entirely different formula from that hypothesized to cause liver injury is still considered "possible" by the DILIN's EO-based CAM (*Table 5*). CB was also evident in case 2 due to the assignment of a higher causality score for OEP than Stacker 3, despite the latter containing green tea extract, while OEP was used for 2 months previously without adverse effect (*Table 2*). CB is also present as the authors indicate the OEP supplement was reformulated in March

of 2013, yet the patient began using OEP in January of 2013, contradicting the authors' argument that this patient consumed the reformulated version (*Table 2*). CB is again noted when considering that the DILIN score for OEP was apparently elevated over Stacker 3 due to the latter's long history of use; however, bias is evident compared to cases 4 and 5 where similarly long periods of use were noted for OEP, which did not result in a decrease in DILIN causality scores. In case 3, CB is evident from the implication of OEP despite evidence of an obvious alternative (i.e., an oral contraceptive) as well as the elevated RUCAM score from the expert as compared to the computer (*Tables 1,2*). In cases 4 and 5, CB is evident from the implication of OEP despite the extremely long latency (i.e., 745 and 720 days, respectively) and the fact that this would also indicate a reformulated version of OEP was not consumed (*Tables 2,3*). In case 6, CB was evident as azithromycin's role was minimized. However, the rise of ALT and symptom onset only after azithromycin use (i.e., OEP had already been used for 4 months at this point) and the nearly perfect match of the published onset and general chronology of azithromycin-induced liver injury with that experienced by case 6 indicate it is a more likely cause (53). Turmeric was also ignored (55,56). In case 7, CB was evident as OEP was implicated despite a lack of temporal association for OEP and a closer temporal association with WBVP as well as obvious indications of an acute exacerbation of a chronic HBV infection.

DM

In cases 3–7, DM is demonstrated as the authors note that causality was assigned at least partially because the product was implicated in causing liver injury.

IC

In cases 1 and 7, IC is present as the mere fact that the patient had used OEP at some point prior to liver injury led to an inclusion despite an improbable time to onset. In case 2, IC is present as a causal role was assumed for OEP because of a temporal relationship, despite evidence contradicting the authors' argument that the patient consumed a reformulated version. In case 4, IC is present as the authors implicated the product due to a temporal relationship yet failed to consider evidence which undermined a causal role (e.g., latency of 745 days, non-implicated version of OEP used and lack of exclusion for viral causes and AIH) (*Tables 2,3,4*). In case 5, IC is present as the authors implicated the product due to a temporal relationship yet failed to consider evidence which undermined a causal role (e.g., latency of 720 days, non-implicated version of OEP used, TMP/SMZ use and lack of viral causes excluded (*Tables 2,3,4*). In case 6, IC is present as the authors implicated the product due to a temporal relationship yet failed to consider evidence which undermined a causal role (e.g., azithromycin, other agents and lack of appropriate viral exclusion).

MAB

In case 1, MAB is demonstrated by the lack of consideration and exclusion of viral causes (*Table 4*). In case 2, MAB is present as AIH, fatty liver and various viral causes were either not considered or ruled out with insufficient information (*Tables 2,3,4*). The patient had several criteria consistent with AIH, yet immunoglobulin G (IgG) and anti-mitochondrial antibodies (AMA) were not assayed. The PI excluded the possibility of AIH because corticosteroids had been given without a positive response (*Table 4*). However, demonstrating overreliance upon heuristics (33), a significant amount of AIH cases do not respond to corticosteroids especially in severe and fulminant forms (57), and a lack of response does not exclude AIH (58). In case 3, MAB is demonstrated by the lack of consideration for the oral contraceptive and various potential viral causes which were not excluded (*Table 4*). In cases 4, 5 and 6 MAB is present as various viral causes were never excluded, while in case 5, TMP/SMZ was excluded via insufficient methodology. In case 7, MAB is present as various viral causes were never excluded, HBV was not properly considered and several other dietary supplements were either not excluded or inadequately considered (*Table 4*).

OB

In case 2, overlap was found for MAB and OB as AIH, fatty liver and various viral causes were either not considered or ruled out with insufficient information (*Tables 2,3,4*). In case 3, OB was present as the PI arbitrarily claimed the patient displayed a "typical phenotype for this agent," when there was no established phenotype and a liver biopsy revealed acute hepatitis with cholestasis. OB was also demonstrated as the authors assumed without any evidence that the patient was using a reformulated product from March of 2013. To the contrary, the information provided

by the patient indicated that she resumed taking the same supplement she had taken for nearly two years (i.e., 2010 to 2012), demonstrating it was not the reformulated version (*Tables 2,3*). In case 4, OB was present as various viral causes were never excluded. In case 5, overlap was found for MAB and OB as various viral causes were never excluded, while TMP/SMZ was excluded via insufficient methodology. In cases 6 and 7, there was overlap for MAB and OB as various viral causes were never excluded, while in case 7, there was overlap for MAB and OB as HBV was not properly considered and several other dietary supplements were either not excluded or inadequately considered (*Table 4*).

PPE

PPE was evident in case 5 as TMP/SMZ was effectively discounted as playing a role in liver injury since the patient had previous self-reported exposure 2–3 times prior without difficulty (*Table 3*). However, even assuming the patient had a correct recollection of previous TMP/SMZ use, cases of TMP/SMZ and other antibiotic-induced liver injury have been documented in patients who had previous exposure without adverse effects (23,53,59-63), demonstrating overreliance upon heuristics (33).

PC

PC was evident in cases 3–7 as the authors assumed the patient was using a reformulated version without any evidence (*Table 3*).

RR

RR was evident in case 5 as TMP/SMZ was discounted due to lack of rash in the patient; however, cases of TMP/SMZ-induced liver injury have frequently presented without rash (60-66), again demonstrating overreliance upon heuristics (33).

UP

The UP bias is present in case 1 as to the use of ibuprofen. While it was used after symptoms had begun, it appears there was no effort to evaluate whether the large doses of ibuprofen (6–9 tablets daily) had been used at any point prior to the emergence of symptoms. In case 2, the UP bias was present as no apparent attempt was made to confirm what formulation of Stacker 3 the patient had been using

and how long the patient had been using it; this is important as Stacker 3 formulations have contained green tea extract and other potentially hepatotoxic substances such as *Garcinia cambogia* (23,67-69). Additionally, despite having a history of alcohol abuse, apparently no attempt was made to confirm a lack of abuse or relapse via serum ethanol measurement. The UP is also present once again as no apparent attempt was made to determine which version of OEP this patient had used. While potentially of questionable value, plasma acetaminophen did not appear to be assessed in this or any other case (70). This is particularly important in cases where a history of drug abuse (particularly opioids), chronic pain or psychiatric illness has been noted as patients may be provide incomplete or incorrect information if abusing or unintentionally overusing medications (71-75) (*Table 3*). In cases 3 and 4, the UP was present as the authors apparently made no attempt to determine which version of OEP the patient had used (*Table 3*). Furthermore, the authors neglect to consider in this case and others, the old formulation of OEP was available alongside the reformulated version through July of 2013 (22). In case 5, the UP was evident by the lack of sufficient history taking in this patient and gathering on the part of the physician. For example, despite the notation that the patient was a poor historian of drug or supplement use, no effort was apparently made to confirm her previous use of TMP/SMZ, her actual dates of use for OEP and which version she consumed. In case 6, the UP was present as no apparent effort was made to confirm which version the patient consumed or what role other medications such as those containing acetaminophen and other dietary supplements may have played. In case 7, the UP was present as no effort was made to determine the ingredients of other dietary supplements that were consumed by the patient or in two cases (i.e., NO Xplode and WBVP), the supplements were ignored altogether (*Table 1*).

Transparency of data is an absolute requirement

While the RUCAM is a preferred CAM, it will not overcome inadequate data inclusion. For example, in case 3, the PI scored RUCAM for OEP was 7 (probable), yet the oral contraceptive used by the patient was never entered into the algorithm (*Tables 1,2,7,8*). Had it been entered, the score for OEP would have become a 5 (possible), exemplifying the fact that the RUCAM can only function properly if all relevant data are collected, analyzed and appraised with systematic and transparent methods (18). In cases 2, 3, 5,

6 and 7 agents which were alternative causes of potential HILI/DILI were either not scored via the RUCAM by investigators or those scores were not reported, while errors in RUCAM calculations were found for all 7 cases, which would have been apparent if the necessary data for RUCAM component scoring had been reported (*Tables 2,3,4*). Inclusion of unreported data resulted in a decrease in RUCAM scores for OEP (*Tables 7,8*), with three cases excluded entirely and the remaining four cases graded as "possible." Additionally, calculated RUCAM scores for alternatively implicated agents were either equivalent to or substantially greater than those for OEP, indicating alternative causes (*Table 9*). Most notably, cases 3 and 5 had alternative agents with RUCAM scores of "probable." Finally, while outside the scope of this review, cases 2 and 5 for example, demonstrate the issues with reliance upon human memory and recall for generating quantitative data (23,76).

Global introspection (expert opinion) or RUCAM for evidence-based medicine?

Recently, authors have clarified what an "EO" consists of versus, "expert evidence" in the realm of evidence-based medicine, with the former being defined as "a view or judgement formed about something, not necessarily based on facts," and the latter as "facts (actual or asserted) intended for use in support of a conclusion" (19). The DILIN method obviously relies upon EO while the RUCAM conversely, asks the observer to take "expert evidence" and apply it to an algorithm, making it the obvious choice when practicing evidence-based medicine (15,34). Algorithms in general have been referred to as "a critical tool for evidence-based medicine" and can explain why the RUCAM is used globally in such a large number of cases as compared to the DILIN method (77,78).

Overreliance upon Heuristics

It is also confirmed that EO in this instance relies too heavily upon heuristics (e.g., cases 2 and 5) which are susceptible to cognitive biases and can ultimately result in the patient receiving insufficient or incorrect treatment due to misdiagnosis (32,33,79). Relatedly, the use of supposed, "phenotypes" to assign causality to a single agent (e.g., case 3) is a non-validated approach prone to obvious overlap for clinical and biochemical variables between agents and likely contradictory data when comparing separate cohorts (4,14).

Is the DILIN method superior to the RUCAM?

Some have claimed the DILIN method is superior to the RUCAM because it results in more "positive" cases and supposedly demonstrated higher inter-observer agreement (14,15). However, unlike the RUCAM, the validity of those "positive" cases has not been assessed (80), while the higher rate may be due to lack of exclusion and cognitive biases (9). Regarding a higher inter-observer agreement (i.e., reliability), the difference for complete agreement (i.e., 19% for RUCAM; 27% for DILIN) was not statistically significant (P=0.08), while generally similar conclusions were indistinguishable at 68.9% and 71.1%, respectively (15). Nonetheless, the RUCAM has a calculated accuracy of 87%, which is of equal importance to reliability (8,80-82). Furthermore, the DILIN method may have slightly greater inter-observer reliability because it considers fewer variables (i.e., fewer non-drug and only a limited number (i.e., 3) of drugs or supplements consumed concurrently) (9,10,15).

Does the expert opinion-based method minimize bias?

The DILIN method relies upon group judgements to reduce bias; however, the ability of group judgements to reduce bias and increase accuracy of judgement is dependent upon a high disagreement in initial judgements and group judgements falling outside the range of individual judgements (83), which are both obviously lacking in the DILIN method where initial judgements had 70% agreement and final committee scores changed little from initial reviewer scores (15). Interestingly, one of various "cognitive debiasing" strategies (i.e., aside from awareness and confrontation of CDRs) to reduce cognitive bias in medicine is the use of an algorithm (e.g., RUCAM) for diagnoses (32). While it can be argued that the RUCAM is also susceptible to certain forms of bias (e.g., AB and DM if an agent has been implicated in a publication) the clearly defined components, weighting of each component, forced consideration of alternative causes and penalties for lack of exclusion are likely why the RUCAM may be able to resist bias as compared to the DILIN method, provided it is used appropriately.

Conclusions

Based upon this analysis and review, it is evident that the DILIN method is susceptible to various forms of bias,

subjectivity, inconsistency and a lack of transparency, which have long been noted to be major issues with the EO-based method. The RUCAM is more objective, transparent and may be less susceptible to bias. Possible improvements to the DILIN method would include systematic and transparent data collection, analysis and appraisal in addition to incorporation of debiasing techniques. However, due to issues of bias inherent to EO-based methods, it should not be employed. The RUCAM is a validated tool that should be used in cases of suspected HILI/DILI as a best practice for evidence-based medicine but it can only function properly if data are collected, reported and assessed objectively and transparently. The RUCAM should be utilized to avoid misdiagnoses, incorrect causality attribution, incorrect/insufficient treatment and to determine epidemiological parameters such as HILI/DILI incidence rates.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editor (Rolf Teschke) for the series "Liver Injury by Herbal Products" published in *Translational Gastroenterology and Hepatology*. The article was sent for external peer review organized by the Guest Editor and the editorial office.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tgh-20-148). The series "Liver Injury by Herbal Products" was commissioned by the editorial office without any funding or sponsorship. CMW reports personal fees from USPLabs LLC., outside the submitted work; and the author previously served as a consultant for USPlabs LLC, the manufacturer of OxyELITE Pro until March of 2019 and is a defendant in litigation related to the product. USPlabs was not involved in the conception, writing, or editing of this article. The views expressed here are solely those of the author.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Kulkarni A, Bhatt A. Causality assessment: A casualty of compensation? Perspect Clin Res 2013;4:196-8.
2. Hutchinson TA, Lane DA. Assessing methods for causality assessment of suspected adverse drug reactions. J Clin Epidemiol 1989;42:5-16.
3. Teschke R, Danan G. Causality Assessment Methods in Drug-Induced Liver Injury. In: Chen M, Will Y. editors. Drug-Induced Liver Toxicity. Methods in Pharmacology and Toxicology. New York: Humana, 2018.
4. Danan G, Teschke R. Roussel Uclaf Causality Assessment Method for Drug-Induced Liver Injury: Present and Future. Front Pharmacol 2019;10:853.
5. Koh Y, Yap CW, Li SC. A quantitative approach of using genetic algorithm in designing a probability scoring system of an adverse drug reaction assessment system. Int J Med Inform 2008;77:421-30.
6. Karch FE, Smith CL, Kerzner B, et al. Adverse drug reactions-a matter of opinion. Clin Pharmacol Ther 1976;19:489-92.
7. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. Drug Saf 2008;31:21-37.
8. Benichou C, Danan G, Flahault A. 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 1993;46:1331-6.
9. Teschke R, Eickhoff A, Schulze J. Drug- and Herb-Induced Liver Injury in Clinical and Translational Hepatology: Causality Assessment Methods, Quo Vadis? J Clin Transl Hepatol 2013;1:59-74.
10. Teschke R, Frenzel C, Schulze J, et al. Herbal hepatotoxicity: challenges and pitfalls of causality assessment methods. World J Gastroenterol 2013;19:2864-82.
11. Danan G, Teschke R. Drug-Induced Liver Injury: Why

is the Roussel Uclaf Causality Assessment Method (RUCAM) Still Used 25 Years After Its Launch? Drug Saf 2018;41:735-743.

12. Danan G, Teschke R. RUCAM in Drug and Herb Induced Liver Injury: The Update. Int J Mol Sci 2016;17:14.

13. Teschke R, Danan G. Drug induced liver injury with analysis of alternative causes as confounding variables. Br J Clin Pharmacol 2018;84:1467-77.

14. Tillmann HL, Barnhart HX, Serrano J, et al. Novel Approaches to Causality Adjudication in Drug-Induced Liver Disease. Curr Hepatol Rep 2018;17:276-82.

15. Rockey DC, Seeff LB, Rochon J, et al. Causality assessment in drug-induced liver injury using a structured expert opinion process: comparison to the Roussel-Uclaf causality assessment method. Hepatology 2010;51:2117-26.

16. CEBM (Centre for Evidence-Based Medicine). 2009. Levels of evidence (March 2009). University of Oxford. Available online: https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

17. Short- and Long-Term Use of Benzodiazepines in Patients with Generalized Anxiety Disorder: A Review of Guidelines. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2014 Jul 28. Table 4, Levels of Evidence and Grading of Recommendations. Available online: https://www.ncbi.nlm.nih.gov/books/NBK254098/table/T4/

18. Shakir SA. Causality and correlation in pharmacovigilance. Stephens' Detection of New Adverse Drug Reactions 2004;5:329-43.

19. Schünemann HJ, Zhang Y, Oxman AD. Distinguishing opinion from evidence in guidelines. BMJ 2019;366:l4606.

20. Heidemann LA, Navarro VJ, Ahmad J, et al. Severe Acute Hepatocellular Injury Attributed to OxyELITE Pro: A Case Series. Dig Dis Sci 2016;61:2741-8.

21. Teschke R, Eickhoff A. Suspected Liver Injury and the Dilemma of Causality. Dig Dis Sci 2017;62:1095-8.

22. Willson C. Dietary supplements containing Aegeline and DMAA (1,3-Dimethylamylamine) and their role in liver injury. Int J Med Res Health Sci 2018;7:10-35.

23. Willson C. An evaluation of case report studies concerning dietary supplements containing Aegeline and DMAA (1,3- Dimethylamylamine). Int J Med Res Health Sci 2018;7:162-73.

24. Gibbons RD. OxyELITE Pro and liver disease: statistical assessment of an apparent association. J Statistical Theory Practice 2018;12.

25. Teschke R. Top-ranking drugs out of 3312 drug-induced liver injury cases evaluated by the Roussel Uclaf Causality Assessment Method. Expert Opin Drug Metab Toxicol 2018;14:1169-87.

26. Teschke R, Wolff A, Eickhoff A, Danan G. Is obesity rather than the dietary supplement used for weight reduction the cause of liver injury? JGH Open 2018;2:152-7.

27. Teschke R, Schwarzenboeck A, Frenzel C, et al. The mystery of the Hawaii liver disease cluster in summer 2013: A pragmatic and clinical approach to solve the problem. Ann Hepatol 2016;15:91-109.

28. Teschke R, Eickhoff A. The Honolulu Liver Disease Cluster at the Medical Center: Its Mysteries and Challenges. Int J Mol Sci 2016;17:476.

29. Teschke R, Eickhoff A, Wolff A, et al. Liver injury from herbs and "dietary supplements": highlights of a literature review from 2015 to 2017. Curr Pharmacol Rep 2018;4:120-31.

30. Sarges P, Steinberg JM, Lewis JH. Drug-Induced Liver Injury: Highlights from a Review of the 2015 Literature. Drug Saf 2016;39:801-21.

31. Walter J, Navarro V, Rossi S. Drug-induced Liver Injury Associated with Weight Loss Supplements. Curr Hepatol Rep 2018;17:245-53.

32. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 2003;78:775-80.

33. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. BMJ 2005;330:781-3.

34. Danan G, Benichou 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.

35. Teschke R, Xuan TD. Suspected herb induced liver injury by green tea extracts: Critical review and case analysis applying RUCAM for causality assessment. Jpn J Gastroenterol Hepatol 2019;1:1-6.

36. Lindgren A, Olsson R. Liver damage from low-dose oral contraceptives. J Intern Med 1993;234:287-92.

37. Elouni B, Ben Salem C, Zamy M, et al. Cytolytic hepatitis possibly related to levonorgestrel/ethinylestradiol oral contraceptive use: 2 case reports. Ann Pharmacother 2010;44:2035-7.

38. Little L, Rule J, Peng L, et al. Herpes Simplex Virus-Associated Acute Liver Failure Often Goes Unrecognized. Hepatology 2019;69:917-9.

39. Li IWS, Chok KSH. Acute hepatitis E virus infection causing acute liver failure requiring living-donor

livertransplantation in a non-pregnant immunocompetent woman. Transpl Infect Dis 2017;19. doi: 10.1111/tid.12699.

40. Linnebur SA, Rapacchietta OC, Vejar M. Hepatotoxicity associated with Chinese skullcap contained in move free advanced dietary supplement: two case reports and review of the literature. Pharmacotherapy 2010;30:750, 258e-262e.

41. Yang L, Aronsohn A, Hart J, et al. Herbal hepatotoxicity from Chinese skullcap: a case report. World J Hepatol 2012;4:231-3.

42. Papafragkakis C, Ona MA, Reddy M, et al. Acute hepatitis after ingestion of a preparation of Chinese skullcap and black catechu for joint pain. Case Reports Hepatol 2016;2016:4356749.

43. Woo HJ, Kim HY, Choi ES, et al. Drug-induced liver injury: A 2-year retrospective study of 1169 hospitalized patients in a single medical center. Phytomedicine 2015;22:1201-5.

44. Melchart D, Hager S, Albrecht S, et al. Herbal Traditional Chinese Medicine and suspected liver injury: A prospective study. World J Hepatol 2017;9:1141-57.

45. Martin DJ, Partridge BJ, Shields W. Hepatotoxicity associated with the dietary supplement N.O.-XPLODE. Ann Intern Med. 2013;159:503-4.

46. LiverTox. National Institutes of Health. Summary of Case 19113. 2019. Available online: https://livertox.niddk.nih.gov/Home/ReferenceCases/oxyelite_pro/19113

47. Kumar M, Jain S, Sharma BC, et al. Differentiating acute hepatitis B from the first episode of symptomatic exacerbation of chronic hepatitis B. Dig Dis Sci 2006;51:594-9.

48. Puri P. Acute exacerbation of chronic hepatitis B: the dilemma of differentiation from acute viral hepatitis B. J Clin Exp Hepatol 2013;3:301-12.

49. Wong VW, Chan HL. Severe acute exacerbation of chronic hepatitis B: a unique presentation of a common disease. J Gastroenterol Hepatol 2009;24:1179-86.

50. Chang ML, Liaw YF. Hepatitis B flares in chronic hepatitis B: pathogenesis, natural course, and management. J Hepatol 2014;61:1407-17.

51. Papatheodoridis GV. Hepatitis B virus treatment: Which patients can have treatment deferred? Clin Liver Dis (Hoboken) 2013;2:15-7.

52. Shah SM, Singh SP. Hepatitis B virus serology: Use and interpretation. Hep B Annual 2007;4:39-54.

53. Martinez MA, Vuppalanchi R, Fontana RJ, et al. Clinical and histologic features of azithromycin-induced liver injury. Clin Gastroenterol Hepatol 2015;13:369-76.e3.

54. Teschke R, Schulze J, Schwarzenboeck A, et al. Herbal hepatotoxicity: suspected cases assessed for alternative causes. Eur J Gastroenterol Hepatol 2013;25:1093-8.

55. Luber RP, Rentsch C, Lontos S, et al. Turmeric Induced Liver Injury: A Report of Two Cases. Case Reports Hepatol 2019;2019:6741213.

56. Imam Z, Khasawneh M, Iftikhar H, et al. Turmeric: An Unlikely Culprit of Drug Induced Liver Injury. Am J Gastroenterol 2018;113:S1658-9.

57. Ichai P, Duclos-Vallée JC, Guettier C, et al. Usefulness of corticosteroids for the treatment of severe and fulminant forms of autoimmunehepatitis. Liver Transpl 2007;13:996-1003.

58. Krawitt EL. Autoimmune hepatitis. N Engl J Med 2006;354:54-66.

59. Faria LC, Resende CC, Couto CA, et al. Severe and prolonged cholestasis caused by trimethoprim-sulfamethoxazole: a case report. Clinics (Sao Paulo) 2009;64:71-4.

60. Dujovne CA, Chan CH, Zimmerman HJ. Sulfonamide hepatic injury. Review of the literature and report of a case due to sulfamethoxazole. N Engl J Med 1967;277:785-8.

61. Berg PA, Daniel PT. Co-trimoxazole-induced hepatic injury--an analysis of cases with hypersensitivity-like reactions. Infection 1987;15 Suppl 5:S259-64.

62. Yang JJ, Huang CH, Liu CE, et al. Multicenter study of trimethoprim/sulfamethoxazole-related hepatotoxicity: incidence and associated factors among HIV-infected patients treated for Pneumocystis jirovecii pneumonia. PLoS One 2014;9:e106141.

63. Steinbrecher UP, Mishkin S. Sulfamethoxazole-induced hepatic injury. Dig Dis Sci 1981;26:756-9.

64. Abusin S, Johnson S. Sulfamethoxazole/Trimethoprim induced liver failure: a case report. Cases J 2008;1:44.

65. Nair SS, Kaplan JM, Levine LH, et al. Trimethoprim-sulfamethoxazole-induced intrahepatic cholestasis. Ann Intern Med 1980;92:511-2.

66. Abi-Mansur P, Ardiaca MC, Allam C, et al. Trimethoprim-sulfamethoxazole induced cholestasis. Am J Gastroenterol 1981;76:356-9.

67. Yu Z, Samavat H, Dostal AM, et al. Effect of Green Tea Supplements on Liver Enzyme Elevation: Results from a Randomized Intervention Study in the United States. Cancer Prev Res (Phila) 2017;10:571-9.

68. Crescioli G, Lombardi N, Bettiol A, et al. Acute liver injury following Garcinia cambogia weight-loss supplementation: case series and literature review. Intern

Emerg Med 2018;13:857-72.

69. Sharma A, Akagi E, Njie A, et al. Acute Hepatitis due to Garcinia Cambogia Extract, an Herbal Weight Loss Supplement. Case Rep Gastrointest Med 2018;2018:9606171.

70. Leventhal TM, Gottfried M, Olson JC, et al. Acetaminophen is Undetectable in Plasma From More Than Half of Patients Believed to Have Acute Liver Failure Due to Overdose. Clin Gastroenterol Hepatol 2019;17:2110-6.

71. Roberts DW, Lee WM, Hinson JA, et al. An Immunoassay to Rapidly Measure Acetaminophen Protein Adducts Accurately Identifies Patients With Acute Liver Injury or Failure. Clin Gastroenterol Hepatol 2017;15:555-62.e3.

72. Lee WM. Acetaminophen and the U.S. Acute Liver Failure Study Group: lowering the risks of hepatic failure. Hepatology 2004;40:6-9.

73. Hovaida M, Torrazza-Perez E, McCarthy D. Liver Function Tests "Gone Viral": Acute Hepatitis of Uncertain Cause. Dig Dis Sci 2018;63:32-5.

74. Frey SM, Wiegand TJ, Green JL, et al. Confirming the Causative Role of Acetaminophen in Indeterminate Acute Liver Failure Using Acetaminophen-Cysteine Adducts. J Med Toxicol 2015;11:218-22.

75. Lane JE, Belson MG, Brown DK, et al. Chronic acetaminophen toxicity: a case report and review of the

literature. J Emerg Med 2002;23:253-6.

76. Archer E, Pavela G, Lavie CJ. The inadmissibility of what we eat in American and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines. Mayo Clin Proc 2015;90:911-26.

77. Paydarfar D, Schwartz WJ. Editorial: an algorithm for discovery. Science 2001;292:13.

78. Teschke R. Idiosyncratic DILI: Analysis of 46,266 Cases Assessed for Causality by RUCAM and Published From 2014 to Early 2019. Front Pharmacol 2019;10:730.

79. Hussain A, Oestreicher J. Clinical decision-making: heuristics and cognitive biases for the ophthalmologist. Surv Ophthalmol 2018;63:119-24.

80. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307.

81. Baratloo A, Hosseini M, Negida A, et al. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. Emerg (Tehran) 2015;3:48-9.

82. Swanson E. Validity, reliability, and the questionable role of psychometrics in plastic surgery. Plast Reconstr Surg Glob Open 2014;2:e161.

83. Sniezek JA, Henry RA. Accuracy and confidence in group judgement. Organ Behav Hum Decis Process 1989;43:1-28.