Jpn. J. Drug Inform., 19(1): 24～31 (2017).

Reliability of the Evaluation Methods Used to Assess a Causal Relationship between Dietary Supplement Intake and Changes in Adverse Events

Mamoru Kitagawa[1,2], Kazuki Ide[1], Yohei Kawasaki[1], Shinjiro Niwata[2], Kumi Matsushita[3], Masayuki Kaji[4], Keizo Umegaki[5] and Hiroshi Yamada*[1]

[1] Department of Drug Evaluation and Informatics, Graduate School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan
[2] Quality Assurance Department, Suntory Wellness Limited, 2-3-3 Daiba, Minato-ku, Tokyo 135-0091, Japan
[3] Kikugawa General Hospital, 1632 Higashi-Yokoji, Kikugawa 439-0022, Japan
[4] Shizuoka City Public Health Center, 24-1 Jotou-cho, Aoi-ku, Shizuoka 420-0846, Japan
[5] National Institute of Health and Nutrition, 23-1-1 Toyama, Shinjuku-ku, Tokyo 162-8636, Japan

(Received October 3, 2016
Accepted February 20, 2017)

Abstract

Objective: This study aimed to confirm whether the methods for assessing the reported causal relationship between dietary supplement intake and adverse events are reliable in the clinical setting.

Design: The relationships between supplement intake and adverse events were assessed using two algorithms proposed in our previous report, and causal relationships were evaluated.

Methods: Twelve raters with a high probability of handling adverse event information examined 200 records of dialogues with supplement users. Each rater independently assessed the causal relationship using the two algorithms. The relationships between supplement intake and adverse events were assessed for all 200 cases. Variability in the evaluation among raters was analyzed for each occupation and the whole group of raters. The distributions of evaluation were analyzed, and inter-rater reliability was evaluated using the intraclass correlation coefficient (ICC) and Fleiss' kappa coefficient.

Results: All events of 200 cases seemed to be slight and within the range of variation in daily life. Almost all cases were classified into two categories as "Possible" and "Lack of Information" by each rater. The ICC values for all raters, pharmacists, dieticians, and health care workers were 0.644, 0.573, 0.678, and 0.694, respectively, and the kappa coefficients using the two algorithms were 0.466, 0.426, 0.468, and 0.519 and 0.481, 0.478, 0.465, and 0.517, respectively. There were moderate levels of agreement based on the kappa coefficients and ICC values.

Conclusion: The two algorithms proposed in our previous report may be reliable in the clinical setting. Their reliability could be enhanced by establishing a unified method of accumulation and recording adverse events for supplement intake, which should be evaluated by more raters using more cases of adverse events.

Key words: dietary supplement, adverse events, algorithm, causal relationship

Introduction

Japan's health food market continues to expand and is said to have reached approximately 16 billion dollars in 2014. Results of a survey conducted by the Consumer Affairs Agency in past two years questioned 10,000 people and found that 60% of consumers used supplements[1].

With the increase in consumption of supplements in Japan for at least the past two decades, some serious health damage due to supplements and the origin of their materials have been reported, such as bronchiolitis obliterans from *Sauropus androgynus*, kidney damage from Aristolochia plants, and liver damage from Chinese diet teas[2-4]. Because of condensing component used to elicit a functional effect in humans, as is one of the characteristics of supplements, it is reasonable to assume that there are some changes in health conditions when using supplements[5]. However, in Japan, there are no established systems for collecting adverse event information for supplements by health care facilities or medical institutions, possibly because it is very

*Corresponding author: Hiroshi Yamada, Department of Drug Evaluation and Informatics, Graduate School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

difficult to recognize those events and evaluate them precisely[6]. Moreover, if we promoted the collection of adverse events for supplements, we have to construct methods to evaluate a causal relationship between supplement intake and the events and their severity.

There are several assessment methods to measure the causal relationships between drug intake and adverse events, such as the Austin Bradford Hill criteria[7] and the criteria provided by the Council on International Organization for Medical Science (CIMOS) working group[8]. These methods have been used both for individual, and accumulated adverse event cases. The Naranjo and Food and Drug Administration (FDA) algorithms were methods that were also used to evaluate adverse events related to drug intake, which were proposed between the 1970s and the 1990s[9,10]. A distinguishing feature of these methods are that they are simple, and therefore easy for the clinicians to use in daily practice. Using the Naranjo and FDA algorithms as a basis, Ide et al. conceived an assessment method for causal relationships that reflects the characteristics of a supplement, and reported high reliability[11]. However, our previous report contained a limited number of raters and adverse events. To apply and widely use our method in the clinical setting, it is necessary to validate our assessment method with different raters and cases.

Here, we examined whether or not our assessment method could be applied to assess the casual relationship between adverse events with supplement intake and health changes. We also determined whether our method could be reliable in clinical settings or not, using both different adverse events, and raters with different occupations.

Methods

1. Study Design

Dieticians, pharmacists, and health care facility workers were used as raters due to their high probability encountering or contacting adverse events information due to intake of supplements. Records of dialogues with supplement users that were directly collected through phone calls and electronic mail sent to customer centers for food manufacturing and sales companies in Japan were used for assessment. Two hundred cases were consecutively extracted from recorded dialogues with users of health food products that contained indications of an unfavorable change in physical condition in 2013. Each rater independently assessed the causal relationship for these 200 cases using the assessment method proposed by Ide et al., and variability in the evaluation among raters was analyzed for each occupation as well as the entire rater population.

This study did not require any approval from the ethics committee because the dialogues contained no personal information.

2. Raters

We recruited raters who wished to participate in this study from Kikugawa General Hospital (Shizuoka, Japan) and Shizuoka City Public Health Center (Shizuoka, Japan) by announcement in advance. We selected dieticians, pharmacists, and health care facility workers as raters in this study.

3. Definition of Adverse Event and Evaluation

We defined adverse event as "an unfavorable change in physical condition," which was recorded in dialogues with the users of dietary supplements. Records of dialogues included unedited words from the supplement users that noted changes in an unfavorable or favorable physical condition and their strength without fixed format. Important medical information, such as the presence/absence of medical drug use or underlying diseases and their types or degrees, were recorded only in the cases of declaration from the users.

4. Evaluation Method and Evaluation Category of Causal Relationship

We used the assessment methods proposed by Ide et al., namely the modified Naranjo scale, and the modified FDA algorithm, which reflect the characteristics of adverse events due to supplements[11]. The methods are shown in **Figs. 1** and **2**. These figures were previously published in a journal (doi: 10.1136/bmjopen-2015-009038) and are licensed under Creative Commons Attribution 4.0 International (CC BY-NC 4.0). A modified Naranjo scale and a modified FDA algorithm were used to assess causal relationships between supplement intake and user change in physical condition in 200 cases.

For both the modified Naranjo scale and modified FDA algorithm, evaluation categories were classified using six levels, including "Highly Probable," "Probable," "Highly Possible," "Possible," "Unlikely," and "Lack of Information." In the modified Naranjo scoring scale, the adverse events were assigned to a probability category using the total score, with ≥ 9, 5–8, 3–4, 1–2, and ≤ 0 corresponding to the respective categories, and a lack of information

No	Question	Yes	No	Do Not Know
1	Are there any notification about the reaction on the label or package insert of the dietary supplement?	+1	0	0
2	Did the adverse event appear after suspected dietary supplement intake?	+2	-1	0
3	Did the adverse reaction improve when the suspected dietary supplement was discontinued?	+2	0	0
4	Did the adverse event reappear when the dietary supplements re-intake?	+3	-1	0
5	Are there alternative causes (other than the dietary supplement) that could on their own have caused the reaction?	-1	+2	0
6	Was the reaction more severe when the dose was increased or less severe when the dose was decreased?	+1	0	0
7	Did the consumer have a similar reaction to the same or similar dietary supplements in any previous exposure?	+1	0	0
8	Was the adverse event confirmed by any objective evidence?	+2	0	0

Fig. 1 Modified Naranjo scale
This figure was published by Ide et al. in 2015. (doi: 10.1136/bmjopen-2015-009038)

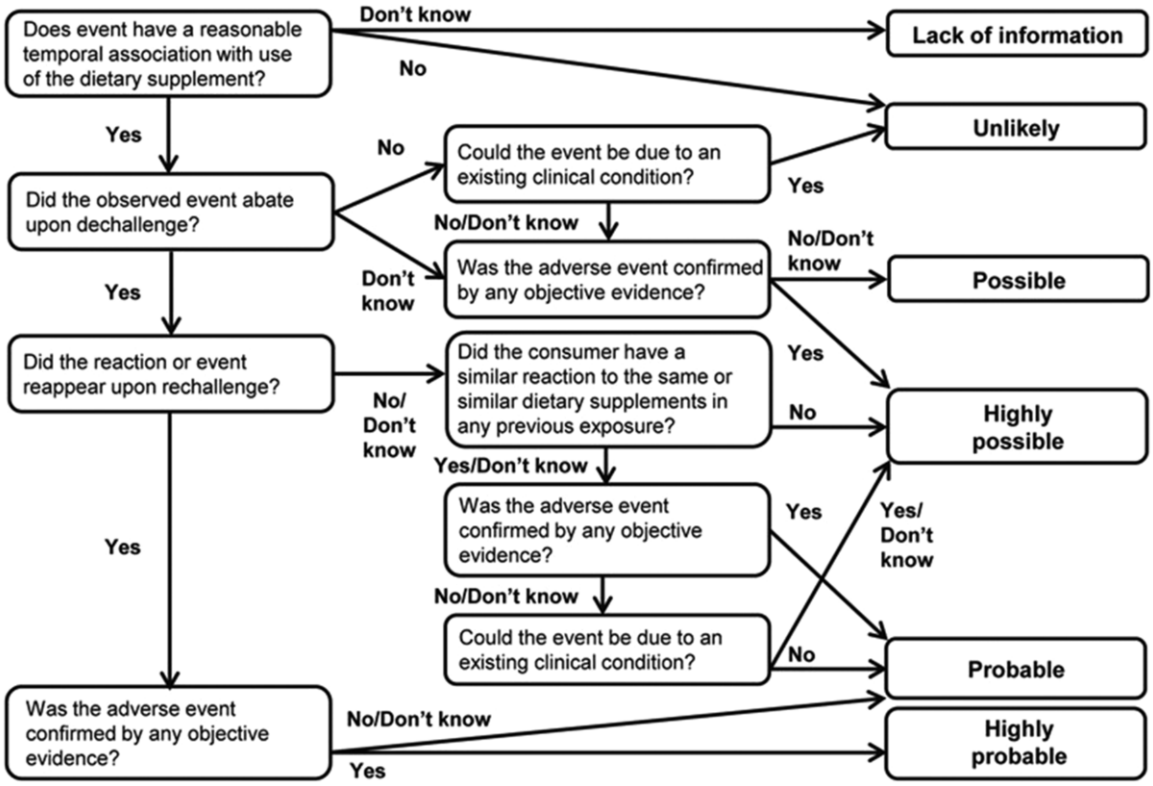

Fig. 2 Modified Food and Drug Administration (FDA) algorithm
This figure was published by Ide et al. in 2015. (doi: 10.1136/bmjopen-2015-009038)

concerning time relationship was categorized as Lack of Information. Twelve raters independently used the modified Naranjo scale and modified FDA algorithm to evaluate the 200 cases.

5. Statistical Analysis

To quantify the agreement level for the modified Naranjo scale, intraclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) were calculated using the methods proposed by Shrout and Fleiss[12]. ICC values were interpreted according to the following criteria: 0.00–0.40, poor agreement; 0.40–0.75, moderate agreement, and >0.75, excellent agreement[13]. Multirater reliabilities for the modified Naranjo scale and the modified FDA algorithm were analyzed using Fleiss' kappa coefficient[14]. The 95% CI of kappa coefficient was calculated from its standard error. Kappa coefficients for each question of the modified Naranjo scale were also calculated. Kappa coefficients were interpreted according to the following criteria[15]: 0.00, no agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–0.99, almost perfect agreement; and 1.00, perfect agreement. All statistical analyses were performed using R version 2.5.13 for Windows (The R Foundation for Statistical Computing, Vienna, Austria) and SAS version 9.4 for Windows (SAS Institute Inc, Cary, North Carolina, USA).

Results

1. Raters

The raters included 4 dieticians (1 man and 3 women), 4 pharmacists (3 men and 1 woman), and 4 health care facility workers (3 men and 1 woman). Each group of raters had differing levels of work experience, with 3–35 years for dieticians, 3–24 years for pharmacists, and 1.5–4 years for health care workers (**Table 1**).

2. Assessment Cases

The 200 cases contained information reported by either users themselves or their family members. The total number of adverse events included 97 events of gastrointestinal changes, including discomfort of the stomach, diarrhea, and constipation, and 82 events related to the skin, including rashes, itching, and redness. With respect to adverse events that did not involve the gastrointestinal tract and skin, there were three reported cases of changes in liver function, one case related to psychoneurosis (insomnia and lethargy), and two cases of an increase in weight and fatigue. All events appeared to be slight and within the range of variation in daily life. Eighteen cases (9.0%) were confirmed using medical records. There were no confirmed cases indicating a relationship with supplement intake by a primary physician.

3. Distribution of Evaluation Category

The average rate of the evaluation category of "Possible" among all of 12 raters was 67.8% (range, 58.0–82.0%) when using the modified Naranjo scale. As for the modified FDA algorithm, the average rate was 69.6% (range, 58.5–80.0%). "Lack of information" was used to classify 19.8% (range, 3.5–31.5%) of cases with the modified Naranjo scale and 19.1% (range, 4.5–31.5%) of cases with the modified FDA algorithm. This distribution was the same regardless of occupation. The proportions of cases classified in these two categories for each job category (dieticians, pharmacists, and health care facility workers) were 85.1, 87.4, and 90.4% using the modified Naranjo scale and 84.6, 90.4, and 91.1% using the modified FDA algorithm (**Fig. 3**).

4. Reliability of Each Method

The ICCs (95% CIs) for each all occupations as well as pharmacists, dieticians, health care workers specifically using the modified Naranjo scale were 0.644 (0.596–0.694), 0.573 (0.506–0.638), 0.678 (0.621–0.732), and 0.694 (0.639–0.745), and the kappa coefficients (95% CI) were 0.466 (0.465–0.467), 0.426 (0.463–0.429), 0.468 (0.464–0.472), and 0.519 (0.516–0.522), respectively. Kappa coefficients for all occupations as well as pharmacists, dieticians, and

Table 1 Rater's occupations and work experience

	Dieticians	Pharmacists	Health care facility workers	Total
Men/women	1/3	3/1	3/1	7/5
Work experience, mean years (range)	23.3 (3–35)	13.8 (3–24)	2.6 (1.5–4)	13.2 (1.5–35)

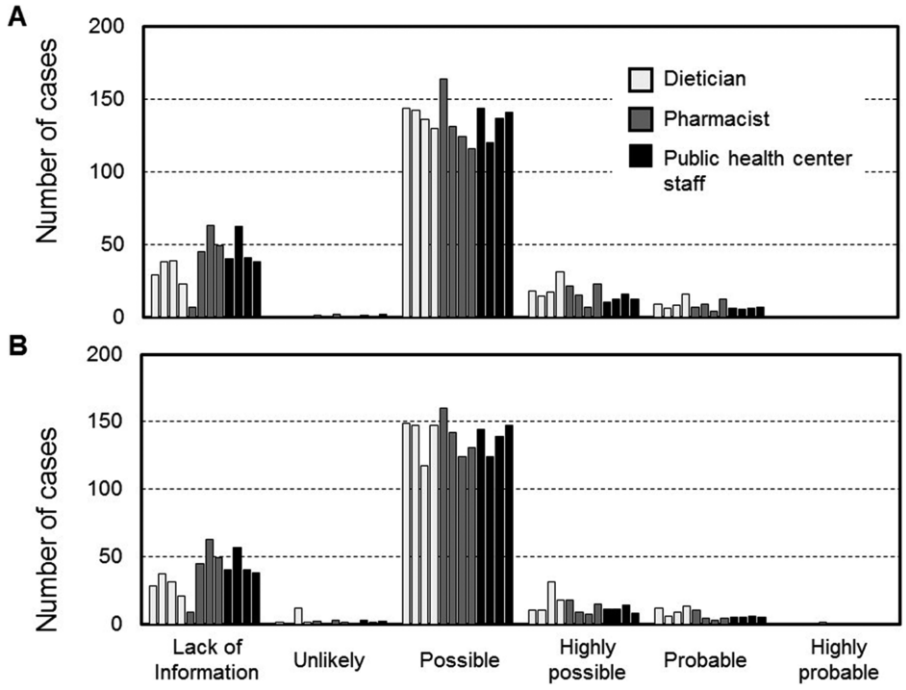

Fig. 3 Classification distribution for a causal relationship between dietary supplements and adverse events
A, Modified Naranjo scale; B, Modified FDA algorithm.

health care workers specifically using the modified FDA algorithm were 0.481 (0.480–0.482), 0.478 (0.474–0.482), 0.465 (0.461–0.469), and 0.517 (0.513–0.520), respectively. The levels of agreement based on the kappa coefficients and ICCs for each method were moderate (Table 2).

Kappa coefficients for each question in the modified Naranjo scoring scale are shown in Table 3. Two questions, "Was there any notification about the reaction on the label or package insert of the dietary supplement ?" and "Was the adverse event confirmed by any objective evidence ?," resulted in low kappa coefficient values.

Discussion

In this study, we evaluated the two methods described by Ide et al., namely the modified Naranjo scale and modified FDA algorithm, which have already been reported for their reliability to assess casual relationships between supplement intake and adverse events, using a different case cohort and different raters. We were able to verify the reliability of both assessment methods using overall kappa coefficients. We also confirmed that since the kappa coefficients for both assessment methods were equal for all occupations and categorized as fair agreement, our assessment methods may be reliable even for different assessment groups and cases.

Table 2 Kappa coefficients and Intraclass correlation coefficients for each occupation

	Dieticians		Pharmacists		Health care facility workers		Total	
	Modified Naranjo	Modified FDA	Modified Naranjo	Modified FDA	Modified Naranjo	Modified FDA	Modified Naranjo	Modified FDA
Kappa coefficients (95% CI)	0.468 (0.464–0.472)	0.465 (0.461–0.469)	0.426 (0.423–0.429)	0.478 (0.474–0.482)	0.519 (0.516–0.522)	0.517 (0.513–0.520)	0.466 (0.465–0.467)	0.481 (0.480–0.482)
ICC (95% CI)	0.678 (0.621–0.732)		0.573 (0.506–0.638)		0.694 (0.639–0.745)		0.644 (0.596–0.694)	

CI, confidence interval; ICC, intraclass correlation coefficients.

Table 3 Kappa coefficients for each question in the modified Naranjo scoring scale

No	Question	Kappa coefficient
1	Are there any notification about the reaction on the label or package insert of the dietary supplement ?	−0.0004
2	Did the adverse event appear after suspected dietary supplement intake ?	0.44
3	Did the adverse reaction improve when the suspected dietary supplement was discontinued ?	0.78
4	Did the adverse event reappear when the dietary supplement re-intake ?	0.57
5	Are there alternative causes (other than the dietary supplement) that could on their own have caused the reaction ?	0.29
6	Was the reaction more severe when the dose was increase or less severe when the dose was decreased ?	0.54
7	Did the consumer have a similar reaction to the same or similar dietary supplements in any previous exposure ?	0.45
8	Was the adverse event confirmed by any objective evidence ?	0.018

However, in order to show the reliability of both assessment methods, additional studies need to be conducted using the same adverse event cases with different raters, and the same raters with different adverse event cases.

Compared to our previous report, this study had a large "Possible" distribution and a low "Lack of information" distribution for both assessment methods. This is thought to be due to the different rater and subject cohorts. In contrast to the 12 dieticians, pharmacists, and health care facility workers in this study, our previous report included 19 pharmacists, dieticians, and medical and pharmaceutical expert raters. In addition, this study also assessed physical changes from supplement users, which rarely are accompanied with treatment information. While no discrepancy was seen in kappa coefficients between both assessment methods for both this study and previous report, the difference in kappa coefficients is thought to be due to differences in the distribution of assessment categories.

Among the separate occupations included in this study, we saw that the kappa coefficients for health care facility workers were somewhat high. However, there was no large discrepancy seen for kappa coefficients between occupations, signifying the versatility of both the modified Naranjo scale and modified FDA algorithm for multiple occupations. Kappa coefficients for health care facility could have also been somewhat high value due to fact that the four workers all worked similar numbers of years compared to pharmacists and dieticians, and there also could have understood and comprehended less due to less experience with cases. Our previous report verified the reliability of casual relationship assessment with 19 pharmacists, dieticians, and experts. However, the degree of conformity between raters was reported as high (excellent), with kappa coefficients of pharmacists using the modified Naranjo scale and modified FDA algorithm being 0.79 and 0.78, respectively[11]. The kappa coefficients of pharmacists in this study using the modified Naranjo scale and modified FDA algorithm were 0.426 and 0.478, respectively, which are considered moderate agreement in terms of assessment of degree of conformity, showing a clear discrepancy with the kappa coefficients reported previously. However, this is thought to possibly be due to a difference in the attributes of the raters. The average work experience for pharmacists in this study was 3.8 ± 7.6 years, while it was 8.6 ± 13.8 years for our previous report, suggesting that even for identical occupations, a disparity in experience and comprehension of adverse events for supplements may contribute to kappa coefficients.

In contrast to previous report, we also assessed the details of dialogues with users that were provided by businesses. This could have affected the kappa coefficients since there is little medical information in reported information from users, and the methods for collecting and recording information are not uniform. The calculated kappa coefficient for each question on the scoring scale, namely "Was there any notification about the reaction on the label or package insert of the dietary supplement ? " and "Was the adverse event confirmed by any objective evidence ? " were low, with values of −0.0004 and 0.0180, respectively. For the question of "Are there alternative causes (other than the dietary supplement) that could on their own have caused the reaction ? ," the kappa coefficient was 0.29, which was low compared to the kappa coefficients of other questions. It shows that raters differentially evaluated the questions.

Information collection methods have not been established public in Japan, and this might affect the kappa coefficient values for some items. If we could collect information concerning adverse events of supplements in structured format, as is done in the US[16], there could be less variation for evaluation among the raters. Similarly, instruction on case evaluation may also be required before making use of the algorithms in clinical practice.

While the collecting and recording methods in Japan for information pertaining to adverse events for supplements is ambiguous and not regulated, establishment of uniform methods would be beneficial. We could not sufficiently assess the validity of this study, which is a limitation. However, an approach to this would be for the same raters to assess more adverse events, or to set a gold standard.

Conclusion

Casual relationship evaluation using the modified Naranjo scale and modified FDA algorithm proposed by Ide et al. are useful with different raters and different cases of adverse events[11]. Though the kappa coefficients in this study were lower than those of our previous report, the reliability of their assessment methods could be enhanced with establishment the unification of collection, recording, and accumulation of the adverse events for supplement intake down a regulation.

Further investigation is necessary using different populations of raters and different cases of adverse events.

Acknowledgments

The authors would like to acknowledge all the raters in this study. We also thank Mr. Yuma Buno for his technical assistance.

Funding: This work was supported by a Health and Labour Sciences Research Grants (Research on Food Safety) from The Ministry of Health, Labour and Welfare, Japan.

Author Contribution: KI, KU, and HY designed this study. MK, SN, KM, and M Kaji performed the experiments. YK and KI analyzed the data. MK and KI wrote the manuscript. All authors read and approved the final manuscript.

Conflict of Interest: MK and SN are the employees of Suntory Wellness Ltd, the company that provided the dialogue records used in this study. Suntory Wellness Ltd. had no role in this study design, analysis, or interpretation of the data, decision to submit the paper for publication.

References

1) Reports of Cabinet Office and Consumer Affairs Agency, Government of Japan, "Survey on the use of health food in Japan. 2012 (in Japanese)," http://www.cao.go.jp/consumer/iinkaikouhyou/2012/houkoku/201205_report.html (accessed 29 August, 2016).

2) Sawahata M, Ogura T, Tagawa A, et al. Sauropus androgynus-associated bronchiolitis obliterans of mother and daughter—autopsy report. *Respir Med CME* 2010; **3**: 214–7.

3) Nishigami E, Kawaguchi Y, Terai C, et al. Progressive interstitial renal fibrosis due to Chinese Herbs in a patient with Calcinosis Raynaud Esophageal Sclerodactyly Telangiectasia (CREST) syndrome. *Intern Med* 2006; **40**: 1059–63.

4) Press Release of Ministry of Health, Labour and Welfare, July 12, 2006 (in Japanese)," http://www.mhlw.go.jp/houdou/2002/07/h0719-3.html (accessed 29 August, 2016).

5) Koike M, Ohtsu F, Sakakibara J, et al. The current status of health food- or supplement-related adverse health effects. *Jpn J Drug Inform* 2013; **14**: 134–43.

6) Umegaki K, Yamada H, Chiba T, et al. Information sources for causality assessment of health problems related to health foods and their usefulness. *Shokuhin Eiseigaku Zasshi* 2013; **54**: 282–9 (in Japanese).

7) Hill AB. The environment and disease: association or causation ?. *Proc R Soc Med* 1965; **58**: 295–300.

8) Council on International Organization for Medical Science (CIOMS), "Management of safety information from clinical trials, report of CIMOS Working Group VI. 2005," http://www.cioms.ch/index.php/publications/printablev3/541/view_bl/66/drug-development-and-use/43/management-of-safety-information-from-clinical-trials-report-of-cioms-working-group-vi (accessed 11 January, 2017).

9) 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.

10) Jones JK. Adverse drug reactions in the community health setting: approaches to recognizing, counseling, and reporting. *Fam Commun Health* 1982; **5**: 58–67.

11) Ide K, Yamada H, Kitagawa M, et al. Methods for estimating causal relationships of adverse events with

dietary supplements. *BMJ Open* 2015; **5**: e009038.
12) Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420–8.
13) Rosner B. *Fundamentals of biostatistics*. Pacific Grove: CA: Duxbury Press, 2005.
14) Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; **76**: 378.
15) Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
16) United States Government Accountability Office, "Report to congressional requesters—Dietary Supplements. (GAO-13-244) 2013," http://www.gao.gov/assets/660/653113.pdf (accessed 29 August, 2016).