Methods for estimating causal relationships of adverse events with dietary supplements

Kazuki Ide, Hiroshi Yamada, Mamoru Kitagawa, Yohei Kawasaki, Yuma Buno, Kumi Matsushita, Masayuki Kaji, Kazuko Fujimoto, Masako Waki, Mitsuyoshi Nakashima, Keizo Umegaki

ABSTRACT

Objective: Dietary supplement use has increased over past decades, resulting in reports of potentially serious adverse events. The aim of this study was to develop optimised methods to evaluate the causal relationships between adverse events and dietary supplements, and to test these methods using case reports.

Design: Causal relationship assessment using prospectively collected data.

Setting and participants: 4 dietary supplement experts, 4 pharmacists and 11 registered dietitians (5 men and 14 women) examined 200 case reports of suspected adverse events using the modified Naranjo scale and the modified Food and Drug Administration (FDA) algorithm.

Primary outcome measures: The distribution of evaluation results was analysed and inter-rater reliability was evaluated for the two modified methods employed using intraclass correlation coefficients (ICC) and Fleiss’ κ.

Results: Using these two methods, most of the 200 case reports were categorised as ‘lack of information’ or ‘possible’ adverse events. Inter-rater reliability among entire assessors ratings for the two modified methods, based on ICC and Fleiss’ κ, were classified as more than substantial (modified Naranjo scale: ICC (95% CI) 0.873 (0.850 to 0.895); Fleiss’ κ (95% CI) 0.615 (0.615 to 0.615)). Modified FDA algorithm: Fleiss’ κ (95% CI) 0.622 (0.622 to 0.622).

Conclusions: These methods may help to assess the causal relationships between adverse events and dietary supplements. By conducting additional studies of these methods in different populations, researchers can expand the possibilities for the application of our methods.

INTRODUCTION

The entire functional food market is estimated to be worth over US$80 billion. This market reached US$32.5 billion in the USA in 2012, with more than half of the adults reporting use of one or more dietary supplements. Sales of dietary supplements have also increased in Japan, with an estimated market size second only to that of the USA. In fact, one study indicated that over 50% of the Japanese population consumes dietary supplements. With the increased use of dietary supplements, a number of adverse events have been reported. Some of these adverse events can lead to severe disability or death, so managing risk and safety is essential in order to protect consumers. Several legal systems have been developed to regulate labelling and manufacturing standards for dietary supplements, but there are no clear systems in place to detect and report adverse events.

Evaluation of the causality of adverse events is essential in order to determine the risk and safety of supplements. It can also help with issue evaluation, signal detection and regulatory updating. Several methods exist for evaluating causality, including the Naranjo scale, the Food and Drug Administration (FDA) algorithm, the Kramer scale, the Liverpool scale, and the modified FDA algorithm. Our methods may be useful for assessing adverse events caused by dietary supplements in clinical settings. This simple and easy method for evaluating causal relationships can contribute to prompt issue evaluation, signal detection and regulatory updating. Additional studies with different populations are needed to expand the possibilities for application of our methods.

Strengths and limitations of this study

- There is no optimised method for evaluating these adverse events.
- We developed two methods for assessing adverse events associated with dietary supplements and inter-rater reliability among entire assessors was classified as more than substantial.
- Our methods may be useful for assessing adverse events caused by dietary supplements in clinical settings.
- This simple and easy method for evaluating causal relationships can contribute to prompt issue evaluation, signal detection and regulatory updating.
- Additional studies with different populations are needed to expand the possibilities for application of our methods.
the WHO scale.\textsuperscript{19} However, these methods are primarily used to assess adverse events associated with medications. They are not optimised for application to dietary supplements. The information available from consumers taking dietary supplements differs from information provided by patients taking medications. Therefore, the development and optimisation of methods to evaluate the causal relationship between adverse events and dietary supplements is essential in order to improve the quality of risk management.

In the present study, we modified the Naranjo scale and the FDA algorithm and then used these to assess 200 case reports of suspected adverse reactions to dietary supplements. The main objective of this study was to test these modified methods using case reports.

**METHODS**

**Study design**

The Naranjo\textsuperscript{12,13} scale and the FDA algorithm\textsuperscript{14–16} were modified for use with dietary supplements. Two hundred case reports were randomly sampled from a database of adverse event reports associated with dietary supplements. Case reports in the database were based on consumers’ voluntary reports through telephone calls to the consumer information centre in Japan and were not standardised for the evaluation of causal relationships. We recruited assessors from six institutions in Japan (University of Shizuoka, Keio University, Kikugawa General Hospital, Shizuoka City Shizuoka Hospital, Shizuoka City Public Health Center and Hamamatsu Institute of Clinical Pharmacology and Therapeutics) by announcement. Nineteen assessors (4 dietary supplement experts, 4 pharmacists and 11 registered dietitians; 5 men and 14 women) enrolled and evaluated the case reports by alternately using the modified Naranjo scale and the modified FDA algorithm. The characteristics of the 19 assessors are shown in table 1. Three dietary supplement experts worked at a general hospital and one worked at a university as a full professor. All four of the pharmacists worked at a general hospital. Four of the registered dietitians worked at a general hospital, and seven worked at a city healthcare centre. None of the assessors received any training in the use of the two scales, and they were not familiar with causal assessment of adverse drug reactions since earlier.

**Assessment scale design**

**Modified Naranjo scale**

The modified Naranjo scale is shown in figure 1. The phrase ‘drug’ in the Naranjo scale was changed to ‘dietary supplement’. The section in question 3 of the Naranjo scale pertaining to a specific antagonist was deleted. Questions regarding placebo and blood (or other fluid) concentrations were excluded. In addition to these changes, the scoring for questions pertaining to readministration and confirmation by objective evidence was changed by adding one point for positive answers to the original version of the Naranjo scale. The adverse event reports were assigned to a probability category using the total scores as follows: ≥9 highly probable, 5–8 probable, 3–4 highly possible, 1–2 possible, ≤0 unlikely. Case reports lacking information about time relationships were excluded and categorised as ‘lack of information’.

**Modified FDA algorithm**

Details of the FDA algorithm have been described previously.\textsuperscript{16} The modified FDA algorithm is shown in figure 2. There was limited information included in the dietary supplement case reports, so the number of options for questions was changed from 2 to 3: ‘Yes’, ‘No’ and ‘Don’t know’. The scale was structured with 4 primary questions and 5 branch questions. The contents of the main questions were as follows: (1) the temporal relationship; (2) changes in symptoms due to the dietary supplement being discontinued; (3) rechallenges and (4) objective evidence from laboratory tests such as a drug-induced lymphocyte stimulation test or patch test. Each of these questions had branch questions relating to: (1) existing clinical conditions; (2) objective evidence from laboratory tests such as a drug-induced lymphocyte stimulation test or patch test and (3) previous experiences of adverse events after taking the same or similar (eg, including the same ingredient) dietary supplement. Adverse event reports were assigned to one of the following probability categories on the basis of the answers to these questions: lack of information, unlikely, possible, highly possible, probable and highly probable.

**Table 1**

<table>
<thead>
<tr>
<th>Table 1 Assessor characteristics</th>
<th>Dietary supplement expert</th>
<th>Pharmacist</th>
<th>Registered dietitian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number, n</td>
<td>4</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Age, mean±SD</td>
<td>65.8±11.5</td>
<td>37.8±7.8</td>
<td>42.2±12.4</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (25)</td>
<td>3 (75)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Women</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td>10 (91)</td>
</tr>
<tr>
<td>Career length, mean years±SD</td>
<td>33.5±2.4</td>
<td>8.6±13.8</td>
<td>13.9±17.4</td>
</tr>
</tbody>
</table>

**Statistical analysis**

In order to quantify the level of agreement in the modified Naranjo scale, intraclass correlation coefficients (ICCs) with a 95% CI were calculated using the methods described by Shrout and Fleiss.\textsuperscript{20} ICCs were interpreted according to the following criteria: <0.40, poor agreement; 0.40–0.75, moderate agreement and >0.75, excellent agreement.\textsuperscript{21} Inter-rater (multirater) reliability for the modified Naranjo scale and the modified FDA algorithm was analysed using Fleiss’ $\kappa$ with a SE.\textsuperscript{22} Fleiss’ $\kappa$ values were also calculated for each question of the modified Naranjo scale.
Fleiss’ \( \kappa \) values were interpreted according to the criteria defined by Landis and Koch\textsuperscript{23}: \(-1.00\), total disagreement; \(0.00\), 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 SAS V9.4 for Windows (SAS Institute Inc, Cary, North Carolina, USA).

**RESULTS**

The modified Naranjo scale and the modified FDA algorithm are shown in figures 1 and 2. All assessors evaluated 200 case reports using the modified Naranjo scale and the modified FDA algorithm. No results were missing from the case report evaluations. The distribution of evaluation results is shown in figure 3A, B. These case reports were based on voluntary consumer reports, included incomplete reporting and were not standardised for the evaluation of causal relationships. Most of the 200 case reports were categorised as ‘lack of information’ or ‘possible’. The median (range) of cases in ‘lack of information’ using the modified Naranjo scale was 64 (8-143) and the corresponding values using the modified FDA scale were 64 (8-142) cases. The ‘possible’ category included a median (range) of 88 (19-136) cases using the modified Naranjo scale and 90 (17-138) cases using the modified FDA scale. The information on dosage, previous similar events and objective evidence was particularly poorly reported in these case reports. A large proportion of the cases were mild. Skin symptoms such as pruritus (n=56) and gastrointestinal symptoms such as abdominal discomfort (n=62) were the most common. However, two serious adverse events related to hepatic dysfunction were reported. In one serious case, a woman started to take a dietary supplement for weight loss. Two weeks after commencing this treatment, her health deteriorated and she presented at a general hospital. Laboratory analyses revealed abnormal hepatic enzyme results and she was diagnosed with liver dysfunction. This condition resolved after over 2 weeks of
hospitalisation. The attending doctor considered that the patient’s dietary supplement had caused her liver dysfunction. In another case, a woman had been taking a dietary supplement for weight control for several months and had experienced fatigue for several weeks. She presented at a general hospital, where laboratory analyses revealed abnormal hepatic enzyme results and she was diagnosed with hepatitis. Her attending doctor considered that this was due to the dietary supplement. The patient’s hepatitis improved after around 2 weeks’ hospitalisation.

Modified Naranjo scale
The ICCs and Fleiss’ $\kappa$ coefficient (Fleiss’ $\kappa$) values for the modified Naranjo scale are shown in Table 2. The ICCs (95% CI) for each assessor group were as follows: dietary supplement experts, 0.865 (0.836 to 0.891); pharmacists, 0.890 (0.865 to 0.911) and registered dietitians, 0.882 (0.859 to 0.903). For the entire group of assessors, this value was 0.873 (0.850 to 0.895). Fleiss’ $\kappa$ values (95% CI) for each assessor group were as follows: dietary supplement experts, 0.598 (0.596 to 0.599); pharmacists, 0.791 (0.790 to 0.792) and registered dietitians, 0.610 (0.609 to 0.610). For the entire group of assessors, this value was 0.615 (0.615 to 0.615). The levels of agreement based on the ICCs for each assessor group and all assessors combined were excellent. Inter-rater (multirater) reliability classifications based on Fleiss’ $\kappa$ were as follows: fair agreement among dietary supplement experts and substantial agreement among pharmacists, registered dietitians and the entire group as a whole.

Fleiss’ $\kappa$ values (95% CI) for each question of the modified Naranjo scale were as follows: item 1 (product labelling), 0.048 (−0.169 to 0.264); item 2 (temporal relationship), 0.530 (0.530 to 0.531); item 3 (changes in adverse event after discontinuation), 0.944 (0.943 to 0.945); item 4 (rechallenges), 0.861 (0.857 to 0.866); item 5 (other factors related to the adverse event), 0.585

Table 2  ICC and Fleiss’ $\kappa$ coefficient values for the modified Naranjo scale and the modified FDA algorithm

<table>
<thead>
<tr>
<th>Assessor Group</th>
<th>Modified Naranjo scale $\kappa$ Coefficient (95% CI)</th>
<th>ICC (95% CI)</th>
<th>Modified FDA algorithm $\kappa$ Coefficient (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary supplement expert (n=4)</td>
<td>0.598 (0.596 to 0.599)</td>
<td>0.865 (0.836 to 0.891)</td>
<td>0.596 (0.594 to 0.598)</td>
</tr>
<tr>
<td>Pharmacist (n=4)</td>
<td>0.791 (0.790 to 0.792)</td>
<td>0.890 (0.865 to 0.911)</td>
<td>0.780 (0.779 to 0.781)</td>
</tr>
<tr>
<td>Registered dietitian (n=11)</td>
<td>0.610 (0.609 to 0.610)</td>
<td>0.882 (0.859 to 0.903)</td>
<td>0.624 (0.623 to 0.624)</td>
</tr>
<tr>
<td>Total (n=19)</td>
<td>0.615 (0.615 to 0.615)</td>
<td>0.873 (0.850 to 0.895)</td>
<td>0.622 (0.622 to 0.622)</td>
</tr>
</tbody>
</table>

FDA, Food and Drug Administration; ICC, intraclass correlation coefficient.
(0.584 to 0.585); item 6 (dose dependency), 0.797
(0.754 to 0.840); item 7 (adverse event history), 0.057
(0.022 to 0.093) and item 8 (objective evidence from
laboratory tests), 0.561 (0.519 to 0.603). Items 1 and 7
showed the two lowest levels of agreement.

Modified FDA algorithm
Fleiss’ \( \kappa \) values for the modified FDA algorithm are shown in
table 2. Fleiss’ \( \kappa \) values (95\% CI) for each assessor
group were as follows: dietary supplement experts, 0.596 (0.594 to
0.598); pharmacists, 0.780 (0.779 to 0.781) and registered
dietitians, 0.624 (0.623 to 0.624). For all 19 assessors, this
value was 0.622 (0.622 to 0.622). Inter-rater (multirater)
reliability based on Fleiss’ \( \kappa \) values were as follows: fair
agreement among dietary supplement experts; substantial
agreement among pharmacists, registered dietitians and
the entire group of assessors as a whole.

DISCUSSION
In this study, we modified the Naranjo scale and the
FDA algorithm and used them to evaluate case reports
of adverse reactions to dietary supplements. These
reports were assessed by dietary supplement experts,
pharmacists and registered dietitians.

Agreement levels for the Naranjo scale, based on ICCs
for each individual group and the assessor group as a
whole, were classified as ‘excellent’. Fleiss’ \( \kappa \) values for each
assessor group and for the group as a whole also
demonstrated more than fair agreement. These results indicated
that the modified Naranjo scale would be useful for evalu-
ating the causal relationships between adverse events and
dietary supplements. It may also have broad utility among
different professions. The only concerns were items 1 and
7 (product labelling and adverse event history, respect-
ively), which produced the two lowest levels of agreement.
To remedy this, assessors might easily obtain the informa-
tion from consumers as they are reporting the adverse
events. Revising these two items and also recording consu-
mers’ reports as they occur may improve the inter-rater
(multirater) reliability and usability of the modified
Naranjo scale.

The modified FDA algorithm showed more than fair
agreement between each assessor group and within the
entire group. Like the Naranjo scale, it has broad utility
and would be useful for assessing the causality of
adverse events.

For both methods, the inter-rater (multirater) reliabil-
ity ratings determined using ICCs and Fleiss’ \( \kappa \) analyses
showed more than substantial agreement in the entire
group of assessors. In fact, the Fleiss’ \( \kappa \) values were very
similar (0.615 for the modified Naranjo scale vs 0.622
for the modified FDA algorithm). Between them, scient-
ists could select the one that best suits their purpose.

A large proportion of the 200 cases assessed in this study
reported mild symptoms, although two serious cases with
hepatic dysfunction were included. Although mild symp-
toms are not life-threatening, they do affect the quality of
life. Therefore, analysis of causal relationships and the pro-
vision of information can improve the safety of dietary sup-
plement usage. The number of serious adverse events was
limited but these can lead to severe disability; the analysis
of causality using this method can lead to prompt diagno-
sis and treatment, as well as regulatory actions.

There were several limitations to this study. The main
limitation was the distribution of evaluation results. For
both evaluation methods, most of the 200 case reports
were categorised as ‘lack of information’ or ‘possible’. This
may reflect the limited information included in the
case reports used in this study. Case reports were based
on voluntary consumer telephone calls and were not
structured to facilitate evaluation of causal relationships.
This might have affected the inter-rater (multirater) reli-
ability ratings. In fact, most of the disagreements among
assessors related to classification as either ‘lack of infor-
mation’ or ‘possible’, while there was fairly good agree-
ment concerning ‘highly possible’, ‘probable’ and ‘highly
probable’ cases. This may be due to the evaluation based
on speculation of each assessor in the cases categorised
as ‘lack of information’ or ‘possible’. Structured or semi-
structured standardised interviews of consumers can
improve the quality of information in case reports. When
designing a structured or semi-structured interview form,
information on dosage, previous similar events and
objective evidence should be requested, in addition to
the essential information regarding temporal association
and discontinuation. Even in the cases categorised as
‘probable’, some of these items of information were
absent. For example, a man started to take a dietary sup-
plement for health enhancement, and then developed
oral inflammation. After discontinuation of the supple-
ment, his oral inflammation resolved. When he started
to take the dietary supplement again, oral inflammation
recurred and he then stopped taking the supplement.
This case included information on temporal association,
discontinuation and rechallenge, but lacked information
on dosage, previous similar events and objective evi-
dence. Validity of the methods may also be a limitation.
We estimated inter-rater (multirater) reliability using
ICCs and Fleiss’ \( \kappa \). However, these methods were not vali-
dated. Future studies could validate these methods in dif-
ferent populations in order to address this limitation and
expand the potential for application of our methods in
other clinical and regulatory settings. For example,
medical institutions and regulatory agencies might use
these modified methods to screen for adverse effects asso-
ciated with dietary supplements, which may accelerate
the detection of harmful events.

The FDA currently operates the Safety Reporting Portal24
for organisations, professionals and consumers. The Safety
Reporting Portal is the electronic version of MedWatch
3500, 3500A and 3500B,25 which are voluntary
reporting forms for adverse events, tailored to
dietary supplements. However, researchers point out
that these data sets contain many incomplete reports.
Other national or local health departments are often

Open Access
the first to detect harm, because these forms are detailed and possibly too complicated for people to use. Combining a screening tool with detailed surveillance will make the reporting system more user friendly. This may promote voluntary reporting and lead to more rapid detection of harmful events.

In summary, we present a modified Naranjo scale and a modified FDA algorithm that may be used to assess the causal relationships between adverse events and dietary supplements. These tools might also be used by regulatory agencies to screen for adverse supplement events, but additional studies are needed to expand the possibilities for application of our methods.

Author affiliations
1Department of Drug Evaluation & Informatics, Graduate School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan
2Kikugawa General Hospital, Shizuoka, Japan
3Shizuoka City Public Health Center, Shizuoka, Japan
4Faculty of Pharmacy, Keio University, Tokyo, Japan
5Shizuoka City Shizuoka Hospital, Shizuoka, Japan
6Hamamatsu Institute of Clinical Pharmacology & Therapeutics, Shizuoka, Japan
7Information Center, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan

Acknowledgements The authors would like to acknowledge the assessors in this study.

Contributors KI, HY and MK designed the study. KI, HY, YB, MK performed the research and collected the data. YK, HY and YK wrote the manuscript. All the authors reviewed and approved the contents of the manuscript.

Funding This work was supported in part by a grant from the Japanese Ministry of Health, Labour and Welfare (No. 24220601 to HY and KI), and a grant from the Japan Society for the Promotion of Science (JSPS) through the Grant-in-Aid for JSPS Fellows (No.15J10190 to KI).

Competing interests None declared.

Ethics approval This study protocol was approved by the ethics committee of the University of Shizuoka (No. 26-6, 2014).

Provenance and peer review Not commissioned; externally peer reviewed.


Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES