To the Editor:

Thank you for the very careful review of our manuscript. We appreciate the interest you have shown in our work. We have considered the thoughtful criticisms and respond below. We have revised the manuscript accordingly. We hope these changes adequately address your concerns.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Criticism: Methods: the major weakness relates to the fact that perceived effects have been measured on a three point scale with anchor points "better", "worse" and "about the same". Strictly this is a categorical scale, not an interval scale. Hence, the scale properties do neither allow performing factor analysis techniques, nor the computation and comparison of mean scores (t-tests and ANOVA) from this categorical scale. The scale properties do allow testing for association, e.g. using chi-square test statistics or using non-parametric tests. This has important implications for the analyses and presentation of findings, which should be given due consideration.

Response: We appreciate the concern of the reviewer about this difficult issue. The standard psychometric approach to scale development, which has been the basis for the validation, use and development of population norms for all currently used psychometric instruments used in medical research, such as the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) and the European Organization for Research in the Treatment of Cancer Quality of Life Questionnaire 30-Item Core Questionnaire (EORTC QLQ C30), treats the Likert response scale as interval, rather than ordinal data, when multiple items are accumulated as summary scales. The summary scale is treated as interval level data. While the assumption is not strictly justified and certainly represents an approximation, the assumption is tested in psychometric analysis, in which the Cronbach alpha value indicates the internal consistency of the responses to the scale items. We certainly concede the theoretical point, but we are reluctant, I hope understandably, to challenge the entire field in the present manuscript. We refer to the following standard texts that endorse our approach, despite its warts.


However, we have altered our Methods section, altering text and adding references to address this caveat.

Criticism: Discussion: second paragraph: appropriate references are needed with the statements that include reference to "previous studies ..." and "Like others ...".

Response: We were referring to the more detailed comments below. We have placed the appropriate
references in the second paragraph as well.

Criticism: Table titles should be made more informative and include also the sample size to which the findings apply, and when relevant, the measurement units; e.g. The title "Table 3. Help from Sen-Sei-Ro" is not informative the way it stands.

Response: We have amended the Titles of Tables 2-5 to be more informative.

Criticism: Table 5: results (e.g. mean values and statistical tests; see comments on Methods)

Response: We believe that the Methods section provides the best place to address these issues, which address the conceptual framework. Tables are problematic sites for extensive text explanations. However, we are of course open to an alternate approach if the reviewers and editor feels more extensive caveats are necessary in the tables.

Criticism: Typos: page 10 "Europe"; page 12 twice "on".

Response: Thank you. We have corrected these oversights.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Criticism: Abstract: conclusions section: change "favorable effects" into "perceived favorable effects"; write CAM in full, this abbreviation may not be common knowledge for all future readers of the abstract.

Response: Corrected.

Criticism: Introduction: change sentence about how long the product has been manufactured and marketed in Japan; change "for the last 12 years" into concrete information such as "since 19xx".

Response: We have made the suggested change, to "since 1991".

Criticism: Results: second paragraph; replace "of these consumers" by "of the respondents", or "of the participants"; The section about positive and negative effects should be revised (in line with the comments made before related to the measurement scale properties).

Response: We have made these corrections as requested.

Criticism: Conclusions: same remark as for abstract: change "favorable effects" into "perceived favorable effects".

Response: Done.

Criticism: Table 4 and perceived effect items: explain the difference in meaning between "maintained or gained weight" and "lost weight"; it is logical that these items load with opposite signs on the same dimension; please clarify.

Response: All individual symptom items were recoded as necessary in order that high scores would consistently indicate greater symptom relief prior to the factor analysis. Hence "maintaining weight" was scored positively, "lost weight" was scored to give a positive value to not losing weight. In this way, we acknowledge that subjects can recognize a difference between avoiding/minimizing weight loss and gaining weight. Moreover, the calculation of the alpha measure of internal consistency requires that all indicator variables have the same valence. We have clarified that issue in the manuscript: "Each item was coded so that increasing values indicated greater relief from symptoms."

Sincerely,

James A. Talcott, MD SM
Corresponding Author