

Author's response to reviews

Title: Test-retest reproducibility of a food frequency questionnaire (FFQ) and estimated effects on disease risk in the Norwegian Women and Cancer Study (NOWAC)

Authors:

Christine L Parr (christine.parr@basalmed.uio.no)

Marit B Veierod (m.b.veierod@medisin.uio.no)

Petter Laake (petter.laake@medisin.uio.no)

Eiliv Lund (eiliv.lund@ism.uit.no)

Anette Hjartaker (anette.hjartaker@medisin.uio.no)

Version: 3 **Date:** 22 December 2005

Author's response to reviews: see over

Dear Editor,

We would like to thank the *Nutrition Journal* for the rapid review, and for constructive and pertinent comments to the previous version of our manuscript. The original comments are included below with a point-by-point answer from the authors, and a description of the changes in the resubmitted manuscript. As can be seen, we received the comments by email with some formatting problems, but we think the points made were clear. We hope you will find the revised version suitable for publication.

1. The association between dietary intake and blood pressure was analyzed cross-sectionally. Presence of blood pressure has strong influences on his or her dietary habit. Most physicians advise patients with high blood pressure to reduce alcohol intake. Dietary intake is not likely disease risk but result of disease. Since the study is undertaken as part of prospective study, the association of dietary intake should be tested with subsequent disease risk. Alternatively the part of β_1 (Bestimated effects on disease risk β_1) should be deleted.

We agree that a prospective analysis would have been more appropriate to investigate the association between blood pressure and alcohol. Our objective, however, was not to investigate this association per se, but to give a practical learning example of the regression calibration method using real data. Numerous studies have been published on the reproducibility of FFQs, but few studies have corrected disease risk estimates based on the results. Thus, we think it is important to illustrate methods for how to investigate potential effects of using data with imperfect reproducibility, and to correct risk estimates for random error. Regression calibration is a convenient method that works with most regression models and study designs. Although the association between alcohol and blood pressure found in our study may not be real, we think the magnitude is representative of the weak diet-disease associations typically seen in nutritional epidemiology.

Not to mislead the readers into interpreting the association between alcohol and blood pressure instead of the effects of the regression calibration, we now emphasize in the manuscript (Abstract/Methods, Methods/Statistical analysis on p. 9, Results/Impact on diet-disease associations (pp. 12.13), and Discussion/Measurement error effects on p. 16) that we just give an example. Table 5 with the results has been reduced to a minimum. Consequently, the parts Results/Impact on diet-disease associations (pp. 12.13) and Discussion/Measurement error effects (p. 16) have been reduced and adapted to the new table.

2. Is the objective of this study to simply assess the test-retest reproducibility of a FFQ, or to compare different statistical measurements to assess reproducibility? Some information such as percentage of misclassification and symmetry of misclassification on the Table 2 is not discussed in the text. Authors should discuss the significance of each measurement in the table, how they differ, and explain why they were all necessary.

In the Background (last paragraph) we now emphasize that the reproducibility is analyzed with different statistical measures, and the reason for this. We do not think that reproducibility can be captured efficiently by one single measure. The different measures complement each other in the interpretation of the reproducibility and will facilitate the comparison with the results of other reproducibility studies.

The percentage of misclassification in ± 1 , ± 2 , and extreme opposite categories are commonly presented in the nutrition literature to assess the degree of misclassification. Under Methods/Statistical analysis (p. 8) we explain that the symmetry of the misclassification indicates if there is a shift towards higher or lower responses in the retest compared to the test. The shifts emphasized in Results/Reproducibility of single food items (pp. 10, 11) are the base for the discussion of seasonal reporting bias under Discussion/Reproducibility of the FFQ (p.13, 2nd paragraph).

3. Readers may not be familiar with the estimation of intraclass correlation coefficients. Notation of ICC(1,1) and ICC(3,1) need be explained in detail.

We interpret this comment as we should describe how the ICCs are calculated. Therefore, we have given the formulae for the coefficients in the text under Methods/Statistical analysis (p. 9), with an explanation of the different variance components in each model, and the main difference between ICC(1,1) and ICC(3,1). The ICCs are a group of several coefficients, so some form of notation is needed, but there is no standard. We used the notation by Shrout & Fleiss where the first number refers to one of three cases of random and fixed effects models used as examples in their paper, which we reference. The second number indicates if the reliability is assessed for single measurements or the mean of several measurements. This is also explained in the manuscript.

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

1. On page 8, line 22, "Normally distributed" should start with lower case.

This has been corrected.

2. On page 11, line 1-2, it is not clear which food items correspond to frequencies with no additional portion size question here. Since this will be an important discussion point, they should be stated clearly within this paragraph.

In the section Methods/The food frequency questionnaire (FFQ) on p. 6, we list the foods with an additional portion size question: "Separate questions about the usual amounts consumed were included for fat on bread, vegetables, fish and fish products, sauces and condiments for fish, meat and meat products, ice cream, chocolate, and cod liver oil supplements". The items with no additional portion size question are the remaining foods in the FFQ, but the list would be very comprehensive. Therefore, we refer to the additional file 1 (Appendix 1) at the bottom of this paragraph (p. 7) for a detailed list of all the food items in the FFQ where we now have specified the items with/without an additional portion size question.

3. On page 16, line 15-16, submitted paper should not be referred.

The part of the discussion where this reference was used has been cut from the paper.

Reviewer 2,

This study dealing with huge NOWAC data is interesting and the manuscript is well-written. However, there are some concerns which should be addressed in the revised manuscript.

1. The authors should carefully and kindly explain the reason why ORs were attenuated in single FFQ when compared with the corrected or calibrated values. The discussion on p16, lines 12-18 seems to be difficult to understand.

Attenuation of effect estimates is the most common result of having measurement error in the exposure, as we now explain more carefully in the Discussion/Measurement error effects (p. 16). In general, the within-person variability or error is largest for single FFQ measurements, smaller for the mean of several measurements, and smallest for the calibrated measurement. With large variability in the exposure data, the association with the outcome will often be attenuated, or weaker than when some of

this variability has been corrected for. The difficult part of the discussion referred to on p. 16 in the previous version has been cut from the new version of the manuscript.

2. The similar part of the Abstract, *Although alcohol intake by regression calibration, is also difficult to understand if the reader has not read the main body of the text. The abstract should be comprehensible without referring other parts of the manuscript.*

Attenuation is a common effect of measurement error in the data. We show that an exposure variable with a relatively high reliability coefficient may still have sufficiently large measurement errors to attenuate risk estimates. We now explain the underlying mechanisms for the attenuation phenomenon more thoroughly in the Discussion/Measurement error effects (p. 16), but we think this is quite detailed and technical for the abstract.