Reviewer's report

Title: Stability of dietary patterns assessed with reduced rank regression; the Zutphen elderly Study

Version: 1 Date: 21 October 2013

Reviewer: Gina Leslie Ambrosini

Reviewer's report:

OVERALL COMMENTS
This is a small cohort study of elderly men in The Netherlands (n=467). The paper aims to describe the stability of dietary patterns in these men after five years of follow up. The authors attempted to identify dietary patterns using reduced rank regression (RRR). The authors' empirical approach is encouraging. However, what could have been a quite straightforward analysis identifying patterns in food intake associated with a set of CVD risk factors, is a quite complicated manuscript. The main purpose of this paper is lost within the complex and numerous methodologies applied and the point of the RRR analysis is almost lost along the way. This manuscript would benefit from simplification.

MAJOR COMPULSORY REVISIONS
1. My major concerns relate to the reasons for, and approach by which, the authors reduced the number of food groups in their RRR pattern analysis. Firstly, it is not clear why the authors decided to apply the ‘Van de Voet’s test’ to identify the optimal number of dietary patterns in their RRR analysis (initially based on 36 food groups). RRR always produces the same number of patterns as there are response variables. Unlike PCA or FA, decisions regarding the number of patterns are not necessary with RRR, but rather, which of the resulting patterns explain a sufficient amount of variation and are most interpretable, to take forward. In this analysis, six patterns would have been identified, by default.

2. The authors report that ‘no dietary pattern based on the initial 36 a-priori defined food groups could be derived to explain the six chosen response variables’. This raises concerns about the appropriateness of RRR using the defined response variables, in this dataset. Do the authors mean that no single pattern explained a sufficient amount of variation in all response variables (and by what criteria)? If so, is this a reasonable expectation, given the number and selected response variables? Or did none of the patterns explain any/much variation? These important points are unclear.

3. In response to the poor fit of the RRR model (or inability to explain variation in response variables), the authors reduced the number of food groups in the RRR pattern analysis. They went about identifying those food groups ‘most associated’ with the response variables, by regressing all food groups onto the individual response variables. Food groups were eliminated using a backward regression...
model. The RRR analysis then attempted to find patterns in the 19 remaining food groups that were associated with the response variables. Line 281 (conclusions): ‘Backwards elimination to select a limited number of food (sub)groups related to the risk factors increased the reproducibility of the procedure and yielded two relatively stable patterns in our dataset’. This is probably not surprising, given that the authors only included food groups in their RRR that had an association with the response variables. I doubt this approach to improve the RRR model is appropriate. Does this potentially circular approach increase the probability of finding similar patterns at both time points? Can the authors assure that this approach does not bias their RRR analysis?

4. The authors refer to their methods as similar to those applied by Schulze et al (ref 22) to obtain a ‘simplified score’. However, Schulze et al did this for a different purpose: to cross validate a dietary pattern from one subset of participants in the Nurses Health Study (NHS) to two other subgroups that did not have response (biomarkers) data. Furthermore, Schulze applied a ‘simplified score’ in the EPIC Study, but AFTER a PCA analysis using all food groups. The simplified dietary pattern scores were simply based on scoring weights from a smaller number of foods (only those loading highest on the pattern). It appears that the authors have done something very different here.

5. Line 256 The authors state ‘Regarding the number of food groups we applied an objective method that allowed the inclusion of the most important food groups for CVD risk factors’ in the RRR analysis. If the most important food groups for CVD are identified using this backward regression method, then the RRR analysis appears redundant. What does RRR bring to this analysis if you have already established which food groups are associated with the response variables of interest (CVD risk factors)? This is unclear. Why are dietary patterns being examined in this study at all? There needs to be a more clear rationale (early in paper).

6. Oddly, the resulting RRR patterns (based on 19 food groups) explain so little of the variation in response variables, again, it is questionable whether these data are suitable for the proposed analysis. Only 8% of the variation in all response variables was explained by the dietary patterns – more than the explained variation in food groups, which is concerning, as the aim of RRR is to explain the variation in response variables, not food groups. 8% is also very low, considering that the authors undertook a complicated process to include food groups that had an association with the response variables in their RRR analysis. Did the authors consider reducing the number of response variables instead of food groups?

7. More detail would be helpful on why the six response variables were chosen together as a response set rather than just individually. For example, what is the rationale for including both total cholesterol and HDL cholesterol as response variables? Uric acid and blood pressure?

8. How were individuals who were taking lipid-lowering drugs or antihypertensives handled in the RRR analysis? How might this impact the RRR
9. Line 64-67 One of the main advantages of RRR is seen as it being a hypothesis testing method i.e. it can be used to test defined pathways between diet and health outcomes (it depends on a-priori information as response variables); it is NOT known as a hypothesis generating method. I am yet to be convinced by the latter viewpoint presented by the authors.

10. Line 72 – the stability or tracking of an RRR dietary pattern has been reported previously by others (see Ambrosini GL et al. Obesity, 2013). The reference (7) by Wosje reported dietary pattern factor loadings, not correlation coefficients (reflecting stability) over time. Furthermore, Wosje et al inappropriately included their outcomes of interest (bone and fat mass) as response variables in their RRR analysis, which would have biased the observed associations between their dietary patterns and outcomes.

MINOR ESSENTIAL REVISIONS

11. Abstract - RRR incorporates both exploratory and a-priori components

12. Line 33 is unclear


14. Abstract - Why are these dietary patterns attractive? Are they associated with disease or health outcomes?

15. Line 59 the RRR method does not ‘belong’ to Hoffman et al. Reduced rank regression has been around as a statistical method for much longer, and applied in other disciplines.

16. Lines 60-68 is confusing to newcomers to RRR and probably not necessary

17. There could be greater emphasis on the fact this is looking at population level stability, not at the individual level, in the most part. This is not well distinguished at present.

18. Was the dietary assessment method validated?

19. Can the authors consider the impact of including fewer food groups in the RRR analysis, in terms of interpretability of the patterns?

20. 143-146 This can be simplified as every individual received a z-score for each of the derived dietary patterns

21. 150 ‘Classify individuals similarly’ needs greater explanation

22. Line 150-152 it is not immediately clear why confirmatory RRR was used, here

23. 159-164 Justification for each sensitivity analysis would be helpful

24. What are the implications of using non fasting lipids as response variables?
25. Why was pattern 3 included if it was difficult to interpret and there was no consistency between this pattern at baseline and follow up? Why was this pattern not ignored, like the other three patterns that have not been described.

26. Line 191 There is not data in the tables (showing fibre or other nutrient intakes) to support this statement.

27. The relatively small sample size and the potential limitations of this on the RRR analysis deserves acknowledgement.

28. The authors could consider the potential limitations of using biomarkers as response variables in their RRR analysis.

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests