Reviewer’s report

Title: The Agewell trial: a pilot randomised controlled trial of a behaviour change intervention to promote healthy ageing and reduce risk of dementia in later life

Version: 3 Date: 7 November 2014

Reviewer: Sebastian Köhler

Reviewer’s report:

Clare and colleagues present pilot RCT data regarding feasibility and acceptability of a health behavior intervention that compares goal setting with and without mentoring to a control condition (general health advice) to visitors of an Agewell center. Primary outcomes are change in mental and physical activity levels. Effect sizes are calculated that are supposed to inform sample size for an efficacy RCT on healthy ageing and dementia prevention. The article is generally well written, though it could be more succinct, and the topic is of interest to the field. The major strength is the assessment of a broad range of secondary outcomes that could inform about further fine-tuning of the intervention. At the same time, the authors should be careful not to base their conclusion on selective findings and avoid over-interpretation, especially given the low sample size and high likelihood of spurious findings. I hope that my comments help in strengthening the manuscript.

Major compulsory revisions

1. The overall conclusion that goal setting (with or without mentoring) might lead to “improvements in cognition, health, diet and physical fitness” is not supported by the data, at least not if stated in such general terms. For instance, effect sizes (ES) for contrasts between control and GS groups for overall cognition were negligible, for delayed recall (a major neuropsychological predictor for AD type dementia) controls improved even more than goal setters! Only the Trail Making Test showed low-to-moderate support. For the other secondary outcomes, a similar mixed pattern is observed of no, more or less benefit in the intervention arms. The conclusion should be much tempered to more accurately reflect the heterogeneity of findings.

2. What effect size (Cohen’s d) did the authors consider meaningful given the aim of the study, i.e. inform sample size calculation of a larger RCT? Some ES are bold in the tables, but it is not clear why. Did they exceed some pre-specified threshold? If so, which as it and why was it chosen?

3. Related to this, why are the ES favouring the control condition not in bold? Table 3: Cultural participation ES = -.27, CVLT delayed recall ES = -.25, MEDAS ES = -.20, Predicted aerobic capacity ES = -.31, Back scratch right ES = -.37, left ES = -.48. This information is different from that of the more negligible ES. Now it gives the impression as if ES favoring the intervention are the only informative data.
4. Depressive symptoms increased in goal setters but not controls. Is there any explanation for this? Is that worrisome given that depression is a major risk factor for dementia, something the authors aim to prevent (see eg Norton et al., The Lancet Neurology 2014)?

5. Analyses were not intention-to-treat. Why not? Five participants dropout in the intervention arms, but none in the control arm, yet no attempt was made to augment missing data (eg by multiple imputation). Could this have biased results and how? Please discuss.

6. I am not a great expert regarding health economics, but would it make sense to base cost-effectiveness on the future efficacy RCT that this study wants to inform instead of a pilot RCT that is notoriously underpowered (as acknowledged by the authors)? Now costs are calculated or 75 participants only.

7. Related to this, what sample size is now needed for such an RCT?

8. Please check whether statistical analyses are appropriate. For chi-square tests, many cells are empty or contain low numbers (<5). Fisher’s exact test might be more appropriate. For t-tests and AN(C)OVA, please make sure you check assumptions, especially regarding approximately normal distributions and homogeneity of variances.

9. The two goal setting conditions (mentoring yes/no) show highly heterogeneous outcomes. I would argue that this suggests they cannot be treated as one group in planned contrasts. Please discuss.

10. I do not understand how participants can be blinded to the intervention as they must know whether or not they set goals and whether or not they were being mentored. If the study was not fully explained, including the type of intervention they might be randomized to, before asking consent, this does not comply with good clinical practice. Please comment.

Minor essential revisions

11. While it is stated that everybody attending the center was invited with "no exclusions" (p.8, l.204), some were excluded at a later stage (eg dementia). Please specify any exclusion criteria.

12. Cohen himself considered a d=0.5 as moderate, so d<0.5 is low - or low-to-moderate at best. This should be changed.

13. Having declared that the study did not aim to study nominal statistical significance, it is somewhat strange to read about significance differences throughout the result section. Looking at Table 1, some major differences exist between groups that do not reach significance, probably due to sample size issues (eg level of education, SES, material deprivation). How meaningful are these comparisons then?

14. Randomization was stratified by gender, yet the GS group contains only 4% male participants as compared to 21% in the GM group. Please clarify if randomisation was successful.

15. Qualitative results are very positive throughout. Could they be motivated by
socially desirability? I wonder whether more critical feedback was also provided. This could speak to further fine-tuning of the intervention. What about the participants who became depressed, any informative feedback from them? Any hints that this was related to the intervention (failing to reach goals? social pressure)?

16. Table 1: “Living situation” the percentage “living alone with others” for the GS group is given as 15.7%. This should be 66.7%, no?

17. Table 3: Please state what it is that is presented in parenthesis in the column “Estimated marginal means”. Are these standard errors (they seem too low for standard deviations)?

Discretionary revisions

18. The introduction is rather lengthy and could be condensed. For instance, several theories are listed that underlie design of the intervention (p.5, l128-133). Maybe this could be moved to the discussion (eg as a strength of the intervention being grounded in behavior change theory).

19. I personally find the presentation of the qualitatively results rather lengthy, too. This could be substantially reduced by stating the overall findings and moving verbatim answers to an appendix for those interested.

20. Introduction, p.3, l.80-81: Very minor, but the study by Norton et al. lists only what they identified as the 7 major risk factors. So, more than 30% might be attributable to the whole (known and unknown) universe of modifiable factors. A qualifier such as "at least one-third" or explicitly stating "by targeting seven major modifiable risk factors" would clarify this.

21. Methods, p.6, l.163-172: repeats much of the aims given already. This could be shortened.

22. Table 1: “Goal attainment”, it is not stated what the numbers represent (number of goals per category achieved). Please add as a footnote.

23. Tables: Please give means, standard deviations and percentages with the same precision throughout, i.e. one decimal, so also for 25.0 instead of 25 (without decimal).

Level of interest: An article whose findings are important to those with closely related research interests

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.