Author’s response to reviews

**Title:** Implementation of a lifestyle intervention for type 2 diabetes prevention in Dutch primary care: opportunities for intervention delivery.

**Authors:**

Paulina WA Vermunt (prediabeteseindhoven@gmail.com)
Ivon EJ Milder (ivon.milder@rivm.nl)
F Wielaard (f.wielaard@chello.nl)
Caroline A Baan (caroline.baan@rivm.nl)
Jos DM Schelfhout (j.schelfhout@chello.nl)
Gert P Westert (g.westert@iq.umcn.nl)
Hans AM van Oers (hans.van.oers@rivm.nl)

**Version:** 4 **Date:** 2 July 2012

**Author's response to reviews:** see over
To the editor,

We would like to thank the editor for the opportunity to resubmit our manuscript. As advised by Mrs. Collins we have consulted a statistician and we have endeavored to make the changes suggested by the editor. Please find editors’ comments and our specific responses below,

1. The reviewer, Claire Collins, is correct that a cluster analysis or hierarchical analysis is appropriate when analyzing patients nested within their clinician. However, this may not make a great difference in this case as the randomization was done at the level of the individual patient, not at the level of clinician or practice. Additionally, their main outcome, dropout rate did not show a difference between groups, so a cluster or hierarchical analysis is not likely make a great difference here. The other statistically differences that were shown were within the intervention group. A discussion of the choice of analysis without clustering would be appropriate in the methods or limitations section.

In concordance with the advice of the editor and Mrs. Collins, our statistician recommended us to take the effect of higher-level clustering into account. We therefore repeated all analyses using a multilevel approach (level 1: individual; level 2: nurse practitioner). As the clustering effects on the GP level (level 3) were neglectable after accounting for the effects of the NP, the GP level was omitted. Although the conclusions did not change when accounting for the multilevel structure of the data, most p-values were slightly altered. We marked the adjusted p-values in the results section of our manuscript in red. Furthermore, we acknowledged the changes in the statistical methods used in line 180-188.

2. However, while the authors have added a sample size discussion, it is not adequate. Sample size calculations usually cite values from pilot studies or known literature and also specify what power the study aimed for and the clinically significant difference on which they based their calculation, in addition to the plan for analysis for the main outcome. I also do not see this sample size calculation in their article that gives the details of the conduct of the study.

The sample size calculation of our study was based on the main outcome diabetes incidence. To detect a difference in diabetes incidence of around 50% (amongst others DPS, DPP and SLIM studies; references 2 and 3 in the manuscript) with a power of 0.8, 82 individuals were needed in each arm. However, as implementation of lifestyle interventions in real life settings is challenging, we expected more modest differences between the two study groups.

At the start of the study, no data on the difference in diabetes incidence between groups in ‘the real world’ were yet available. Following Cohen’s conventions (reference 25 in the manuscript), we therefore used an effect size of 0.1 in our power calculations (corresponding to a small sized effect). In these calculations, 393 individuals were needed in each arm. When a post-hoc correction for correlation on the nurse practitioner level was applied (variance 0.03) this number changed to 405.

As in total 925 individuals could be included, this allowed for a dropout rate of approximately 15%, which was in line with others (DPS, DPP: reference 2). We have added a section on the power calculations in our study in line 169-178 of the methods section.
3. Alternatively, the authors should report the power of their study to detect a difference, since their main outcome, dropout rate, was not significantly different between groups. When the reader sees "no difference" in a randomized trial, their next question should be "Did the trial have sufficient power to show a difference if one exists?" The authors have not satisfied the reader on this question.

In this study no difference in dropout-rates between the two study groups was observed. Accounting for the individuals who were lost to follow-up because they developed type 2 diabetes, the statistical power to detect small, but clinically relevant differences in dropout-rates between the study groups (Cohen’s conventional effect size of 0.1) was 0.835. It is therefore unlikely that the lack of a difference in dropout-rate between the groups is explained by a lack of statistical power. We have commented on the statistical power to detect a difference in dropout-rate in line 303-308.