Reviewer's report

Title: Effect of lifestyle intervention for (pre)diabetics in real-world primary care: propensity score analysis.

Version: 2 Date: 11 June 2011

Reviewer: Magnolia Cardona-Morrell

Reviewer's report:

REVIEWER’S COMMENTS

This is a very important topic and the paper could contribute to enhance the evidence for or against translation of results from RCTs into real-life interventions. Negative and neutral results are needed in the printed literature to minimise publication bias in diabetes prevention studies. The language and references adhere to reporting standards.

The sample is smallish but the methods of a retrospective cohort with matching propensity scores are relatively sound. It is regrettable that the evaluation was retrospective rather than prospective, given the national scope of the intervention. However, limitations of data and methods are well acknowledged in the discussion. The statistical methods need reviewing or further explaining. Subject to these, the conclusions may be appropriate. It is hard to judge in the current form whether conclusions are supported by the data.

Major compulsory revisions

The research question is not specified clearly in the abstract or the study design section of the manuscript. It is implied but should be explicit.

The study type should be called by its name in the abstract and methods: a retrospective cohort with a control group matched by propensity score.

Eligibility

Summarise reasons why 2,362 people who ended up in the control group were not eligible for referral. Was it the 'motivation' factor or the co-morbidity factor? or was it social or geographic reasons?

In Figure 1. specify N= for prediabetics and T2D in both the intervention and control group nodes of the flow-chart. This distinction is important.

Outcome measures p.8

The authors mention that exercise levels were monitored during the quarterly checkups. Yet on page 13 of the discussion they argue that they had no data on treatment adherence after the first consultation. Analysis of level of exposure may give important clues to the success factors for interventions in real life (we know they are determinants in intensive, highly controlled settings where the RCTs were conducted). Was there similar information on people NOT referred to
the program? Perhaps authors could attempt to analyse these data or alternatively explain the reasons for omission in the limitations of the study, beyond assuming that “effect was inherent to the intervention”.

Results & statistical analyses

It is unclear if the number of matched control was the same for each person in the intervention group. Also unclear is whether matching on all those variables was done for each individual or whether some individuals were matched on some variables and other individuals on others. This may explain why the matching was not completely balanced (table 1). Given this differentials in age, presence of COPD or CVD, ad marital status, the analysis needs to be either linear regression (for continuous outcomes) adjusting for the above variables among others or conditional regression analysis (for dichotomous outcomes).

The column ‘Adjusted effect intervention (95%CI)” in Tables 2 and 3 doesn’t seem to be showing Relative Risks or Odds Ratios. The estimate of the effect could also have been a difference or a ratio between intervention/control. It is unclear what the estimate is and why the p value is treated as “adjusted” if there was an imbalance on age, CVD, COPD and marital status at baseline.

It appears that the estimates shown in the “intervention” and “control” columns are the mean BMI, FPG, BP, etc.

I may have misunderstood the above from the ‘statistical Analyses’ section. But in general, the reader would benefit from a more comprehensive explanation of the statistical methods on page 9, and self-explanatory notes under tables 2 and 3.

Discussion

While the BeweegKuur is a national intervention, the evaluation 186 participants from 10 practices is hardly a representative nation-wide sample of all the regions in the Netherlands. Make the over-statement (second sentence in the discussion) reflect that you evaluated a regional or sub-sample or pilot rather than imply you conducted a nationwide evaluation of translation in PHC.

If possible, in order to improve the power of the study, expand the sample analysed by including data from more than 10 PHC centres using the same inclusion criteria and MPS analysis. Results may be very different.

Alternatively, I would strongly suggest the authors revise analysis for the diabetics only (i.e. excluding pre-diabetics). Their sample of pre-diabetics is too small for meaningful interpretation and the predictors of success for pre-diabetics may be different from the predictor of success for people with a diagnosis. Combining both groups may have diluted the true effect of the intervention.

Discuss the possibility that lack of effect was due to impact evaluation conducted too early (first few participants enrolled at 1 Jan 2008, lifestyle counsellors were inexperienced with the first few participants, access to services was not sorted out, etc ) and sample size may have been insufficiently powered.

Minor essential revisions
Emphasise the extensive matching on propensity scores efforts because this is one of the strengths of the study, but clarify the intervention/control matching ratio.

Also clarify if the use of participants with outcome information 1 year before and after the intervention is the reason for the different N= in tables 2&3 for each adjusted model.

Discretionary revisions

Perhaps suggest in the title that this is a preliminary evaluation or a pilot evaluation (larger sample could come later and show better results)

Perhaps under the limitations, consider addressing the fact that the BeweegKuur underwent formative and process evaluation but data were not prospectively collected for an impact evaluation as it was designed as a pragmatic program to be run in routine PHC.

It was interesting that despite not observing changes in BMI or exercise level between intervention and control groups, but there were differences in FPG and HbA1c. Any possible explanation for this? Adherence to dietary recommendations? The DPP in China and India showed that diabetes prevention is possible without weight loss.

In the conclusions section authors mention that process evaluation 'might reveal the barriers and facilitators...' As I understand it, BeweegKuur already conducted process evaluation and this information is available [your Helmink reference 2010]. Perhaps include some of the old recommendations from the 2008 process evaluation & recommend that these be explored further.

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'