Reviewer's report

Title: Predicting complete loss to follow-up after a health-education program: number of absences and face-to-face contact with a researcher

Version: 3 Date: 20 July 2011

Reviewer: Vicki Kristman

Reviewer's report:

The research question and hypotheses are important and well defined. However, I do not believe that the authors have adequately addressed some of the concerns of the previous review. Additionally, the methodology used for some new analyses is of concern.

Major Compulsory Revisions

1. Discussion, Limitations paragraph (second last paragraph): Exposure measurement validity is not something that you can just state as a limitation in the discussion. If there is no evidence for a measure being able to measure what it purports to measure at all then that exposure measure should not be used in the study. Therefore, all reference to attention to health and interest in research as an exposure measure in this study should be removed, including the hypothesis as there is no valid way to test that hypothesis with the data collected in this study.

2. Results, second paragraph: The results regarding diagnosis are confusing and poorly described. Additionally I have concerns regarding the methodology used to obtain these results. “Not having a diagnosis of connective tissue disease was a … predictor of complete loss to follow-up.” Does this mean that all those with the other diagnoses except connective tissue disease were more likely to be lost to follow-up? It would be clearer to say that “Those having a diagnosis of connective tissue disease were less likely to drop out than those not diagnosed with connective tissue disease”. Additionally, there are two other severe problems with the diagnosis analysis:

2a) How did you account for the non-independence in the diagnoses analyses? As indicated in table 1; 409 people had 479 diagnoses and therefore diagnoses cannot be treated as an independent variable – this violates the assumption of independence in logistic regression.

2b) In table 2, I am curious as to why you only presented results for two of the eleven diagnoses listed in Table 1. This is a form of publication bias. What was the rationale behind presenting depression and connective tissue disease but no others?

3. Abstract and Table 2: P-values should be removed as these are particularly sensitive to sample size. Show the odds ratios and 95% confidence intervals instead – this will also help to address the concerns one reviewer had about
study power.

Minor Essential Revisions

4. Background, second paragraph, first line: First sentence in second paragraph is not accurate. Inflating the target sample size to compensate for expected loss will minimize problems of imprecision, but may not have any effect on bias; in fact, selecting participants with certain characteristics could introduce selection bias into your study if those characteristics are associated with either the exposure or outcome under study.

**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.