Reviewer's report

Title: A cluster randomised controlled trial of the effectiveness of and lessons learned from a multifaceted, interdisciplinary educational intervention for general practitioners providing primary care to patients with systolic heart failure

Version: 2 Date: 12 November 2008

Reviewer: Allan Donner

Reviewer's report:

This article is essentially a “Lessons Learned “ paper that reports on a largely failed trial(failed in the sense that it yielded an inconclusive result on the primary endpoint after the investment of considerable resources). On the plus side ,the authors provide a very thoughtful and detailed discussion of the possible reasons for this failure which should be helpful to future researchers planning a new study in this field.

Major comments:

1. An important strength of the study is that all eligible practices and patients were identified prior to randomization. This avoided problems of selection bias that are well-known to arise when patients are recruited in an opportunist and unblinded fashion after their practice has been randomized.

2. In discussing the strengths and limitations of their study on pages 14 and 18, the authors state their belief that the internal validity of the trial was high and further implying that the inconclusive findings were largely due to problems of external validity. However I think this is somewhat overstated since some of the problems they describe (eg, inability to blind GPS to treatment group, poor timing of the intervention, inability to deliver it correctly, etc.) are clearly those that threaten the internal validity of the trial.

3. The use of significance tests to compare baseline characteristics is poor practice in any randomized trial. This is because any observed difference can only be the result of chance, thus negating the main purpose of performing a significance test. The problem is compounded in cluster randomized trials that compare individual-level characteristics without taking into account the clustering as then the resulting p-values will be biased downwards. Thus the “significant” p-values reported at the bottom of Table 2 could well be non-significant if appropriate adjustments were made to the t-tests and chi-square tests that produced these p-values.

4. It is stated in the Study Conclusions that the TTT intervention led to improved provider performance. However this conclusion should be regarded with some caution as it results from the lone statistical comparison (out of about 30) that results in a p-value less than 0.05. Thus chance alone might also be a plausible explanation given the large number of comparisons performed.
5. The discussion on sample size calculations is incomplete.
   (i) What is the justification for choosing 6.6 as the size of intervention effect to be detected?
   (ii) The statement that the expected ICC value was based on a “recommendation” is not very helpful without further elaboration.
   (iii) What is meant by “serial correlation” in this context? Are the authors referring to the correlation between the baseline and final measurements? If so, how was this taken into account in the sample size calculation?
   (iv) It should be specified as to whether the sample size calculation was one- or two-sided.

6. Given that only 37 GP’s participated out of 750 GP’s that were initially approached, can any comparisons be made between the baseline characteristics of the participating and non-participating physicians? This might help in understanding to what extent the study conclusions can be reasonably generalized (or not).

Minor comments:

Page 8
1. The primary outcome is not stated correctly. It is more accurately defined as the as the quality of life score measured at the 7 month follow-up point.

2. Page 14
   In line 3, “neither “ should be “either”

3. Page 15
   In line -7, the phrase “ultimately relativises the under-recruitment” is unclear.

4. Page 17,
   In line -7, “selected” should be “highly selected”.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

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

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