Author's response to reviews

Title: A heart failure self-management program for patients of all literacy levels: A randomized, controlled trial [ISRCTN11535170]

Authors:

Darren A DeWalt (dewaltd@med.unc.edu)
Robert M Malone (rmalone@med.unc.edu)
Mary E Bryant (bbryant@med.unc.edu)
Margaret C Kosnar (margaret_corley@med.unc.edu)
Kelley E Corr (kelley_corr@med.unc.edu)
Russell L Rothman (russell.rothman@vanderbilt.edu)
Carla A Sueta (paris@med.unc.edu)
Michael P Pignone (pignone@med.unc.edu)

Version: 3 Date: 9 December 2005

Author's response to reviews: see over
December 8, 2005

BioMed Central Editorial Team  
BioMed Central Health Services Research  
Middlesex House  
34-42 Cleveland Street  
London W1T 4LB  
UK

Dear Editorial Team,

With this letter I am pleased to resubmit the manuscript: A heart failure self-management program for patients of all literacy levels: A randomized, controlled trial [ISRCTN11535170]. We appreciate the thoughtful reviews by Drs. McAlister and Doughty and we have made most of the changes that they suggested including providing additional data in the manuscript. I hope our responses to the reviews and the modifications in the manuscript clarify any ambiguities regarding our study. We have aimed for transparency so the readers can understand the intervention we performed and the outcomes we observed.

The following pages include the reviewers’ comments in italics followed by our responses in regular text.

I feel the BMC HSR review process has yielded a stronger manuscript. I look forward to your comments regarding this revision.

Sincerely,

Darren A. DeWalt, MD, MPH
Responses to Reviewer Comments

**Reviewer: Finlay A McAlister**

**Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)**

1. This paper would be substantially strengthened if the authors had data on causes for hospital readmissions and could provide the data for "CHF readmissions" or even "cardiac readmissions" and "CHF deaths", as well as the all-cause readmissions/death data they do report.

   In our study, we categorized admissions as cardiac or non-cardiac. Cardiac admissions included heart failure (edema, shortness of breath), but also other cardiac causes such as chest pain, uncontrolled atrial fibrillation, etc. We have added this data to the paper. We did not classify cause of death and cannot comment on CHF deaths. The method for cardiac readmission data collection and classification is on page 9 and the results on page 14.

   1a. Also, I'd suggest they report the hospitalization data as # of pts hospitalized at least once/# of pts in that arm of the trial as well as the way they chose to report it (total # of hospitalizations/# of pts in that arm of the trial).

   We have added this information on page 14.

2. The glaring imbalance in Table 1 is the substantially poorer MLWHF scores in the control group at baseline (12 points higher and p=0.0028). The remaining imbalances described in para 3 on page 13 are minor (and many were not statistically significant) and I'd suggest shortening that paragraph substantially. I am surprised that adjustment for baseline imbalances increased the size of the treatment effect in favour of the intervention given that the intervention group had better QOL and a trend towards less severe NYHA class distribution, as well as being more likely to be taking ACEi/ARB and digoxin (granted controls were more likely to be on B-blockers)- as such, I would have expected adjustment for these positive prognostic factors to reduce the apparent impact of the intervention.

   Because of differences in opinion between reviewers, we have elected to not shorten this paragraph but have clarified the results as suggested by reviewer 2.

   2a. What variables were included in the adjustment? I would suggest these be explicitly defined in para 4 on page 13 (I know the investigators described which variables they would consider for adjusting in the methods section, but I'd like to know which ones were actually included in the adjustments).

   We have done so. The variables included in the adjustment are now on page 14 in the paragraph recommended (the paragraph was on page 13 in the previous version).

3. Given that they modified the MLWHF questionnaire from the usual version, do they have any data they can provide confirming the reliability/validity of their modified MLWHF questionnaire? Did they develop the self-efficacy and knowledge questionnaires de novo for this project? Any reliability/validity data for these scales? Can the authors provide some guidance in the text for readers such as myself un-familiar with these
scales what a clinically important difference would be (ie. what does the 12% improvement in knowledge scores and 2 point improvement in self-efficacy scores described on page 14 mean?). Given the questions over the modified scales performance characteristics in this patient population and since neither the patients nor the research assistants performing the questionnaires with the patients and collecting the data were blinded (as per para 2 on page 18), my enthusiasm for these subjective endpoints is muted.

We do not have validity or reliability data on the self-efficacy or knowledge scales as they were created specifically for this study. It is customary to develop self-efficacy scales for an individual study because self-efficacy is measured regarding the specific behaviors the intervention is attempting to change. More general self-efficacy measures do not usually contain enough specificity to give insight into the effects of a behavioral intervention. Knowledge scales are similarly specific to the knowledge associated with the educational intervention. These endpoints are not as important as the clinical endpoints of hospitalization or quality of life. They are more important as process variables to demonstrate that the intervention occurred and that the patient had exposure to the teaching and motivation. Had we not seen any difference in knowledge and self-efficacy, we may have been concerned about the fidelity with which the intervention was delivered. We agree with the reviewer that these endpoints should not be the source of great enthusiasm and have refrained from embellishing them in the discussion.

We have recognized the limitations for research participants with low literacy of many current quality of life instruments and attempted to improve on one of the limitations by changing the response scale from a well known instrument (Minnesota Living with Heart Failure Questionnaire). We do not know the reliability or validity of the instrument after these changes and the lack of effect on this outcome by the intervention could be related to measurement error as we discussed in the limitations.

4. Given how small the "inadequate literacy" subgroup is and the imbalances at baseline between controls and interventions within that subgroup, I'd suggest deleting the first 2 paras on page 15 and merely saying that "there was no statistically significant effect modification between literacy and the intervention". As this subgroup is even more under-powered than the main trial comparison, I think it is a stretch to say that this disease management intervention in CHF has a greater effect in patients with low literacy (as implied currently on page 17 of the manuscript).

We have softened our conclusions with respect to the inadequate literacy sub-group but we would prefer to retain the analysis to show the results for this sub-group.

5. It would be helpful to provide demographic details on the total patient population with heart failure cared for in their GIM Practice (to ensure that the 118 patients deemed eligible for this trial were somewhat representative of the 500 patients with CHF in their practice). Alternately, I would suggest that the authors provide details on what the reasons were for exclusion in the 367 patients with CHF who were deemed ineligible for this study.

Figure 2 lists the reasons for ineligibility and the numbers of patients deemed ineligible based on each reason. We have chosen to not add this information to the text because not all readers will want to review this list. However, if the editors prefer it included in the text of the article, we can certainly do so.
6. It is difficult to isolate the impact of particular elements of their intervention (the 1 hour educational session + educational booklet + 10 telephone calls over 6 months + scheduling of appointments with their attending physician for patients experiencing worsening symptoms) and, given the apparent intensity of the resources needed, did the authors collect cost data and can the authors make some comment in their discussion on this point?

We did not collect cost or resource use data for this study - we have added a paragraph to the discussion to elucidate these points on page 19.

7. I suggest that the authors clarify how diagnoses of heart failure were made. I'm not clear from the description at the bottom of page 5 whether a patient taking lasix for peripheral edema could be included in this study without objective assessment of their LVEF or confirmation of CHF diagnosis by at least 2 observers or 1 observer using standardized criteria?

Yes, such a patient could have been included in the study. We only required a clinical diagnosis of HF by the treating physician, as determined by direct interview of the provider, AND taking loop diuretics, AND NYHA class II-IV symptoms, AND one of the following: 1) chest x-ray findings consistent with heart failure, or 2) ejection fraction <40% by any method, or 3) a history of peripheral edema. Although objective assessment of LVEF was not required, all patients in the study did have objective assessment, and the results are indicated in Table 1 under the category of systolic dysfunction.

8. How was “moderate to severe dementia” defined and did they adjust standard dementia rating scales (such as the Folstein MMSE) for literacy level?

We did not perform MMSE – our determination of dementia was based on the primary care physician’s clinical judgment. We have added this to the methods section.

9. Who collected the outcome data and who ascertained clinical endpoints? I take it that neither the study staff nor the patients were blinded to treatment allocation, but were the individuals reviewing the admissions and deaths blinded to treatment allocation?

The initial collection of the data by our RA was not blinded, nor was their collection of outcome data which is noted in the discussion; however, the adjudication of reason for admission was performed by the lead author (D.D.), who was blinded to intervention status. We have noted the limitation of non-blinded outcome assessment in the Methods and discussion.

10. From my reading of page 10, it appears as if the investigators relied solely on patient self-report for collecting baseline data - is this just a typo? (I imagine they must have used the clinical charts and medical records as well).

Our principal method of data collection was by patient interview. In some cases, this was supplemented by chart review, but not in a standardized manner. We have added text to the methods to reflect this method.
Discretionary Revisions (which the author can choose to ignore)

11. Why the upper age limit on the eligibility criteria? I was surprised how young their CHF patients appeared to be (the patients in our Heart Function Clinic have an average age of over 70 years).

We chose the upper age limit because we were concerned initially that too many patients would be eligible. This was incorrect and needs to be changed for future studies. Of note, our heart failure population may be somewhat younger than those found in some previous studies, due in part to the prevalence and severity of hypertension as a causative etiology.

12. I'm surprised they designed this study and calculated sample size to detect a 9 point difference in MLWHF scores - this is nearly twice as large as the effect seen in the ACEi vs. placebo trials in CHF patients. (not relevant to this manuscript, but for future studies they may be planning)

The anticipated effect size was based on the difference observed in our pilot study, and was also limited by total resources available. The reviewer is correct that even a smaller difference may be important and worthy of detecting. However, we found no difference.

13. While the authors statement at the bottom of page 16 that self-management training programs reduce hospitalization rates to a similar extent as comprehensive programs employing multidisciplinary teams and clinics is technically correct, it should also be acknowledged that multidisciplinary clinics/teams have been shown to reduce mortality (RR 0.75, 95% CI 0.59-0.96 in the meta-analysis in ref #6) while self-management programs have not (RR 1.14, 95% CI 0.67-1.94).

We have added this point.

Reviewer: Robert N Doughty

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. See comments below regarding the adjustment in the results for the differences in baseline characteristics. I think that this may overstate the results and the authors should tone down this result and acknowledge that these results may be overstating the results just based on this one study.

Minor Essential Revisions

1a. The primary end points are described as combined death/all-cause hospital admission and quality of life (with Minnesota Living With HF Questionnaire, that was modified for people of low literacy). These outcome measures were assessed at 6 and 12 months: which time interval was the primary outcome time measure? (minor essential revision)

12 months- we have noted this in the text on page 9.
2. A major point is the statistical presentation of the primary outcomes with adjustment according to differences in baseline characteristics. The method of randomisation was by concealed random number generation and with the proposed sample size of 150 small differences in baseline characteristics may have occurred but should have been due to chance. From Table 1 the only meaningful (and statistically significant) baseline differences were that in the control group there were:
   a. fewer men,
   b. more receiving beta-blocker therapy, and
   c. patients had a higher mean QoL score.

However, the authors have described multiple factors associated with higher hospitalization rates from prior studies and then adjusted for these if the relationship between the variable and the study group or outcome was <0.3. With this analysis the odds ratio for the outcome of death or readmission changes from 0.69 (non-significant) to 0.53 (significant). While the pros and cons of this sort of statistical adjustment can be argued the issue that I see is that there were not marked clinical differences in the baseline characteristics to start with. This is a fundamentally important part of this manuscript as whether the study holds a positive or negative result hinges on this analysis. My sense is that this sort of programme may well result in important reductions in clinical events, but I am concerned that with this statistical adjustment that the true magnitude of these benefits is over emphasized.

Our technique for adjustment is grounded in sound epidemiological methods as described in Hernandez et al., [1] and we believe it to be appropriate, but we acknowledge that adjustment is not universally accepted. It is for that reason that we presented both adjusted and unadjusted results. Particularly, prespecified adjustment based on a priori hypotheses was shown to be more statistically valid than adjusting based only on those factors that were statistically different at baseline. [1] While the unadjusted results are not statistically significant at the p < 0.05 level, we believe the difference in magnitude of the point estimate between the adjusted and unadjusted results are rather small, and support the conclusion that it is most likely that the program is effective in reducing admission by 30% or more. These findings are in accord with previous studies of similar interventions. Additionally, by using negative binomial regression, we have chosen a very conservative analysis strategy. Other studies have used Poisson regression [2] by which our results would have been statistically significant before and after adjustment. However, based on the likelihood ratio test of negative binomial vs. Poisson, the most correct approach is to use negative binomial. This reflects the fact that multiple hospitalizations for an individual are not completely independent events.

3. Further to the above discussion of the statistical adjustment, the authors should also bear in mind that the original sample size was 150 subjects, but the study was stopped after 127 due to funding constraints. The primary end point for which the study was powered was the QoL score which was a negative result based on the sample recruited. A comment to this effect should be made. (minor essential revision)

Comment added on page 17.

4. Some follow up data on the use of medication including the ACE inhibitors, and beta-blockers would be valuable to understand both the results as well as how this form of education impacted on patient adherence to medical therapy. (minor essential revision)

We do not have data on the use of these medications at the end of the study. We agree with the reviewer that this information could be useful. However, our intervention did not
address the prescribing behavior of the physician, so we would expect that any changes made over time would have been similar in both groups.

5. Page 13, para 3, there were fewer men in the control group, not fewer women. While there were slight differences between the groups at baseline with regard to hypertension, diabetes (2 patients only difference), ACEi/ARB use, % with systolic dysfunction, these differences were small and did not reach statistical significance; this should be reflected in the text and clarified in the results section as it is misleading as written. (minor essential revision)

We have clarified as the reviewer suggested.

References
