Author's response to reviews

Title: Characteristics of control group participants who increased their physical activity in a cluster-randomized lifestyle intervention trial

Authors:

Lauren A Waters (l.waters1@uq.edu.au)
Marina M Reeves (m.reeves@sph.uq.edu.au)
Brianna S Fjeldsoe (b.fjeldsoe@sph.uq.edu.au)
Elizabeth G Eakin (e.eakin@sph.uq.edu.au)

Version: 2 Date: 10 December 2010

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

Thank you for the opportunity to respond to reviewer comments on our manuscript (1350393895395144), entitled “Characteristics of control group participants who increased their physical activity in a cluster-randomized lifestyle intervention trial.” Below, we detail our response to each comment, with corresponding changes highlighted in the manuscript.

We believe that the changes made in response to the reviewers’ comments have improved the manuscript, and hope that the editorial team will now find it suitable for publication in BMC Public Health.

Regards,

Lauren Waters
PhD Student
Cancer Prevention Research Centre
The University of Queensland, School of Population Health
Reviewer 1

Minor Essential Revisions

1. While claims that the modest effect sizes in physical activity interventions trials are due in part to improvements in the control conditions is technically correct, the public health relevance of this statement is dependent on what this change is attributable to. The authors suggest that it is the effect of the 'minimal' intervention received. The authors therefore present, as the basis of the study, that the identification of sociodemographic, medical and health behaviour characteristics associated with this change could identify sub-groups of the population whom interventions of this intensity are sufficiently supportive to encourage physical activity. If the change, however, were due to research reactivity effects I would suggest that the implications are that the research methods used in research of this kind could be consistently concealing or under estimating therapeutic benefits of interventions. It is not possible from this study to make this distinction. Nonetheless these issues should be acknowledged throughout both the introduction and the discussion of the research findings.

We agree that there are a number of potential explanations for control group improvements in physical activity intervention trials. Our recent systematic review of control group changes in physical activity intervention trials (under review) showed that there were several factors associated with control group improvements, including: the number of assessments participants undergo, the mode of measurement administration, the exclusion of participants already meeting physical activity guidelines, pre-existing health status and body mass index. Three of these factors relate to the characteristics of participants enrolled in the trial, and it is this finding upon which the research question addressed in the present study is founded. A paragraph has been added to the introduction to acknowledge the numerous hypotheses for control group improvements (lines 77-90).

“A number of potential explanations for control group improvements have been posited, including: the Hawthorne effect (where participants improve in the experimental variable being tested due to awareness of being observed); social desirability bias (the propensity to report behaviour that is compatible with social norms); regression to the mean (a problem associated with intra-participant variation and measurement error, which may occur in trials using pre- and post-intervention measurements, particularly when behavioural screening is employed to select an inactive sample); the effects of measurement (when measurement is sufficient to produce a change in behaviour in the absence of a formal intervention); or, the recruitment of a highly motivated volunteer sample.”
It is also possible that control group improvements could be attributable to the minimal level of intervention often delivered to control group participants in physical activity intervention trials (e.g., brief advice, standardised print materials).”

As suggested by Reviewer 1, we now explicitly acknowledge in the Discussion that the impact of undergoing measurement cannot be distinguished from the effect of the brief intervention received by the usual care group (lines 318-320).

“It should be noted that the impact of undergoing assessments cannot be distinguished from the effects of receiving the brief intervention.”

2. Socio-demographic factors represent relatively crude prognostic factors. More proximal determinants of behavioural change such as behavioural intentions, self efficacy, and other behavioural and cognitive factors are likely to provide more reliable and sensitive predictors of behavioural change as they are consistent with behaviour change theory. While this is acknowledged by the authors in the manuscript, it remains an important limitation of the study.

This limitation is acknowledged in the discussion (lines 383 - 388)

“As psychosocial and cognitive variables were not measured in this trial, we were not able to assess the capacity of these characteristics to predict control group improvements. While assessing these characteristics would require a more intensive screening process, some research suggests that they are potentially more useful in predicting changes in physical activity than socio-demographic variables [48]. A broader range of predictive variables may be necessary to more accurately predict successful behaviour change.”

3. Changes in the control condition (and associations with this) could simply represent a statistical artefact of regression to the mean. I suspect that participants recruited to the trial (being from a primary care setting and with chronic disease risks) may have had quite low levels of reported physical activity at baseline which one would expect could only naturally improve over time. This possibility should be addressed in the manuscript.

Regression to the mean has been acknowledged as a possible explanation for improvements in participants’ physical activity on lines 81 - 83:

“...regression to the mean (a problem associated with intra-participant variation and measurement error, which may occur in trials using pre- and post-intervention measurements, particularly when behavioural screening is employed to select an inactive sample) [20-22]”
However, regression to the mean is likely to have a greater impact on control group change in trials that screen to exclude participants with higher baseline levels of physical activity (as participants who would have regressed down towards the mean on re-measurement are excluded). Screening for baseline physical activity was not conducted in the Logan Healthy Living Program and we therefore feel that regression to the mean would not have contributed greatly to the observed control group improvements. Further, while it is possible that control group participants in our study would have improved over time given their low baseline activity levels, it is equally possible that these participants (particularly given the high rate of co-morbidities) may be more resistant to change, or less likely to demonstrate natural improvements in physical activity.

4. Were participants blind to group allocation? If not, changes in the control condition (and the associations you find with this change) could represent reactivity to the experimental situation (see Shaddish, Cook and Campbell). Participants not receiving the ‘treatment’ may be motivated to show that they can do as well as those who do receive the treatment and compensate accordingly.

Due to the cluster-randomised design, participants in the usual care group were not aware that there was a more intensive intervention arm, therefore we do not believe that randomization to the ‘control group’ would have been a motivator for change. This has been acknowledged in the manuscript (lines 155-156):

“Participants in the UC group were not aware that there was a more intensive treatment arm.”

The potential impact of reactivity has been addressed in the Introduction on lines 77-90, as noted in the response to comment #1.

5. Changes in reported behaviour of controls (and associations with this) may also arise due to the repeated assessment (Kypri et al, Prev Med 2005;761-766) and should be stressed as being part of the ‘minimal’ intervention received by the control groups in the discussion.

We agree with Reviewer 1 and also consider the brief intervention and measurement to be an inclusive package of treatment delivered to the usual care group. The following has been added to the Discussion (lines 318 - 323) to highlight this:

“It should be noted that the impact of undergoing assessments cannot be distinguished from the effects of receiving the brief intervention. However, measurement itself could be considered to be, and used as, an intervention. Measurement has been shown to impact on participants’ physical activity in the intervention trial context. Brief behavioural assessments have been shown to...
successfully reduce hazardous drinking, and the same may be true for physical activity.”

6. The authors rightfully suggest that more research is required to identify those whom may respond to brief interventions. I would suggest that this is best done through experimental research. In light of the considerations above, we may need novel research methods and designs to examine this issue, including the use of, ‘no intervention’ comparison groups, blinding to group allocation or ‘post test only’ outcome assessments to discern the effect of what is received by participants from other research reactivity effects.

We agree with the reviewer’s comments and have cited a relevant study in the Discussion on line 322. The positive impact of measurement on physical activity in the intervention trial context has been demonstrated by van Sluijs and colleagues (Journal of Clinical Epidemiology, 59, 404-411, 2006) using a Solomon-4 group research design. The measurement employed in this trial was, however, extensive and unlikely to be practical outside of the context of evaluating an intervention. Regardless, van Sluijs and colleagues’ novel approach to empirically assessing the impact of measurement paves the way for future research.

7. The focus of the manuscript, both of the introduction and discussion is largely on predictive factors of control group participants. The value in comparisons with predictors of change in the intervention group should be clarified, or removed.

Tests to examine interactions between a common set of predictor variables and group and found that the associations between some predictor variables and meaningful improvement in physical activity differed significantly by group. We therefore stratified analyses by group. We conducted the analysis in the TC group to descriptively compare whether the predictors for the UC group were unique to that group or generic to involvement in the study overall. Clarification of the importance of using an analysis of TC group predictors as a comparator for results observed in relation to the UC group is provided on lines 107-114, as below:

“To our knowledge, no previous study has investigated whether certain individual characteristics are predictive of physical activity improvements specifically for control group participants. The purpose of the current study is to identify baseline demographic, health and behavioural characteristics that were associated with a meaningful physical activity improvement among control group participants in the Logan Healthy Living Program. In order to provide an appropriate point of comparison for these results, we discuss them in relation to characteristics that were independently predictive of improvement among participants in the intervention group.”
We also note in the Discussion that a number of previous studies have examined the predictors of physical activity improvement for the intervention group only, or for the pooled control and intervention sample. By providing a descriptive comparison of the factors that respectively influence the usual care and intervention groups, we aimed to illustrate that this approach may not be appropriate because it is possible that predictors of improvement are dependent on the intensity of the intervention received. We address this on lines 326 - 332.

“Previous studies that have looked at the predictors of physical activity change have either used data from the intervention group only or pooled data from intervention and control groups. The results of the current study suggest that this latter approach may be inappropriate, given that predictors of successful physical activity improvement appear to differ for the control and intervention groups. This may be because predictors of physical activity improvement in response to an intervention are dependent on the intensity of that intervention.”

For further discussion on our decision to conduct independent analyses for each randomized arm, see our response to comment 8, below.

**Reviewer 2**

**Major compulsory revisions:**

8. It was unclear whether the aim of this paper is to find a parsimonious model to predict change of #60 minutes or a priori list of independent factors associated with the change. If a parsimonious model is the objective, it is limited by the variables collected. If the objective is to identify significant factors (with the constraint of variables collected) between the two randomised arms, then the variables included/adjusted in the model have to be comparable. The model selection process of using p=0.2 level for retaining variables in the model need to be justified. Also multicollinearity of variables such as education and income need to be investigated.

The main purpose of the study was not to find a parsimonious model to predict a meaningful improvement in physical activity, but to identify the baseline characteristics that were associated with such change for the usual care (UC) group, and separately (for comparative purposes) for the telephone counseling (TC) group. We have made the changes in the manuscript to clarify that the purpose of the study is to determine factors associated with improvement, rather than to find a parsimonious model predictive of change (lines 107 – 114); see response to comment #7, above).

However, we appreciate the reviewer’s comments about the need to examine a comparable set of independent variables. We initially conducted analyses to examine interactions between a common set of predictor variables and group and found that for some predictor variables, the association between the predictor and meaningful
improvement in physical activity differed significantly by group. The fact that there were interactions between some predictors and group, combined with the fact that we had conceptual reasons (consistent with our research question) for conducting separate analyses for each group, supports our use of stratified analyses. We conducted analyses for the control group participants only and then replicated this analysis for the TC group participants, using the same process and same initial list of independent variables, which resulted in a different set of predictors in each final model. We more clearly explain the rationale for the use of stratified models in the manuscript on lines 220 – 223.

“Analyses were stratified by group because there were significant differences between UC and TC groups in the associations of meaningful change with some predictors (baseline physical activity, education; p for interaction <0.05).”

The estimates we present from the models are sufficiently comparable, even though the models don’t contain the same covariates, because each model is adjusted for anything that is related to the outcome, from the same list of baseline characteristics. We know from the backward elimination process that we have not omitted important variables (statistically, the full and reduced models are the same) and with a generous criterion for entry we are unlikely to have missed important variables due to suppression. Estimates from each model are not confounded (within the limitations of the data that we collected) and there is no need for further adjustments.

The model selection process of using p=0.2 level for retaining variables in the model is somewhat arbitrary and is founded in convention (Bendel and Affifi, 1977, Journal of the American Statistical Association, 72, 46-53). It is a more conservative cut point than p=0.05 and allows for the inclusion of independent variables that may not be significantly associated with the dependent variable at the p=0.05 level in bivariate analyses due to confounding.

Multicollinearity of variables included in each of the original models (before the backwards step-wise elimination process was commenced) was previously tested but not reported on. No evidence of multicollinearity was identified. For both the TC and UC groups, the Pearson correlation coefficients for variables included in the initial logistic regression models were under 0.7, tolerance was > 0.10 and the variance inflation factor (VIF) was <10. Confirmation that this investigation was performed, and the absence of multicollinearity, can now be found in the manuscript (line 232 - 233).

“Models showed no evidence of multicollinearity”

9. The authors have examined the differences in predictors between the completers and drop-outs (n=319 vs n=115) for the whole sample. But stratified analyses were carried out (UC and TC groups were modeled separately), have you checked if the completers and drop-outs within randomised arm are comparable?
We thank the reviewer for raising an important point. In response to this suggestion we conducted analyses to examine the differences in baseline characteristics between completers and drop outs by group. Some statistically significant differences between included and excluded participants in the UC and TC groups were found. Paragraphs highlighting this finding have been added to the Statistical Methods and Results sections (lines 212 – 215; 240 - 248):

“There was some evidence of selective dropout, therefore a sensitivity analysis using all randomised participants (n = 206 UC; n = 228 TC) and assuming no change from baseline for those who dropped out of the study, was conducted to examine whether results were robust to the selective drop out.”

“Within the UC group, participants who were excluded from the main analyses (n=40) differed from the included sample (n=166) with respect to income and BMI, with more excluded UC participants reporting ‘don’t know’ or refusing to answer questions about household income (30.0% [n=12] of excluded vs. 13.3% [n=22] of included, p=0.043), and being in the obese BMI category (62.5% [n=25] of excluded vs. 39.8% [n=66] of included, p = 0.008). Within the TC group, participants who were excluded from the main analyses (n=53) differed from included participants (n=175) with respect to smoking status with more excluded TC participants being current smokers (26.4% [n=14] of excluded vs. 8.0% [n=14] of included, p=0.001).”

As a result of these findings, we completed a sensitivity analyses (as suggested by the reviewer) in order to test whether the results of our analyses were unduly influenced by selective drop out. Further details of the sensitivity analyses are included in our response to comment #18 below. The potential implications of the differences between included and excluded participants are raised in the Discussion on lines 341 - 352 (see comment #18).

10. The randomisation process is to ensure patient characteristics are balanced for the two arms. With completed case analysis these may not be true. The sample size is small for a dichotomous outcome study. Though only one variable (smoking) is statistically significant different between the two randomised arms, few influential physical activity variables are on borderline significant such as gender, education. These variables are preferable be adjusted and retained in the model for comparison.

The purpose of this study was not to evaluate outcomes by group, using a dichotomous variable. Rather, as outlined in the response to comment #8 (above), the purpose was to examine separately, the factors that were associated with meaningful change in physical activity in UC and TC group participants.

We did not include variables that were significantly different (or close to being so) between the UC and TC groups because each group was treated as an independent
sample. That is, we conducted analyses for the UC group, and then replicated the same process, using the same initial set of independent variables, for the TC group to demonstrate that improvement in the two groups is potentially associated with different factors. In response to suggestions by reviewer 2, we conducted a sensitivity analysis comparing the results of the completed case analysis with the results observed when all enrolled participants are included (with baseline values carried forward for participants who dropped out of the study). For more information on the sensitivity analysis, see comment #18 (below).

11. Nagelkerke R square is a pseudo-R-square statistics. It has severe limitation and cannot be interpreted as in ordinary least squares in explaining the % variance explained in the model where the dependent variable is a dichotomized cut-off. My suggestion for a predictive logistic model is the concordant index or “area under the curve” for model performance and Hosmer-Lemeshow statistic for goodness-of-fit test.

We appreciate the reviewer’s comments on the use of the Nagelkerke R Square statistic and we have omitted the use of this statistic from our interpretation. The Hosmer-Lemeshow statistic for goodness of fit is not recommended with sample sizes < 400 (Hosmer and Lemeshow (2000), *Applied logistic regression*). Given the potential inaccuracies introduced by reporting a statistical interpretation of the fit of the model, we have decided not to report this. However, in the discussion we do raise the point that the predictive capacity of the model may be limited by the non-exhaustive selection of variables included in the model (lines 383-388).

“As psychosocial and cognitive variables were not measured in this trial, we were not able to assess the capacity of these characteristics to predict control group improvements. While assessing these characteristics would require a more intensive screening process, some research suggests that they are potentially more useful in predicting changes in physical activity than socio-demographic variables [48]. A broader range of predictive variables may be necessary to more accurately predict successful behaviour change.”

12. The appropriate method to analyse randomised pre and post data is ANCOVA [see reference]. Analysis of change score is subject to bias and regression to the mean. There is a limit to change depending on your pre physical activity level.

We agree with these statements, however because the purpose of this study was not to evaluate outcomes by group over time, and because our dependent outcome was categorical classification of making a clinically meaningful 60-minute change in physical activity, an ANVOCA was not appropriate to address the research question (see responses to comments #8 and #10). Given the confusion, an effort has been made to further clarify the purpose of the study (lines 107-114).
“To our knowledge, no previous study has investigated whether certain individual characteristics are predictive of physical activity improvements specifically for control group participants. Predictors of control group improvement may be different to predictors of improvement in intervention group participants, as the mechanisms underlying control group change are poorly understood. The purpose of the current study is to identify baseline demographic, health and behavioural characteristics that were associated with a meaningful physical activity improvement among control group participants in the Logan Healthy Living Program. We further compare these to the predictors of improvement in the intervention group.”

13. The study design employed a cluster randomized procedure, that GP practice is the primary sampling unit. There is no adjustment for design effect in the analysis. Was the design effect negligible?

The intracluster correlation (ICC) coefficients for general practitioner practice are reported for each primary outcome in a previous paper (Eakin et al., 2009, *American Journal of Preventive Medicine* 36, 142-149). The ICC coefficient for change in minutes of moderate to vigorous physical activity per week was 0.0000, indicating that there was no correlation of responses within a cluster. Therefore, we have not adjusted for design effect in this study. A sentence addressing this issue has been added to the Statistical methods section on lines 233 - 234:

“We did not correct analyses for clustering as there was no evidence of clustering in physical activity outcomes in the LHLP trial.”

14. Different physical/medical conditions were reported for the two models. Have the authors consider using a composite categorical variable of number conditions? The number with “cancer” was small, hence the large confidence interval.

A composite variable (total number of chronic conditions) was tested and found not to be associated with a meaningful improvement in physical activity (in either group) at the p<0.20 level. In response to the reviewer’s feedback, we have eliminated the individual chronic conditions from the models. This action had a negligible impact on the study findings.

Minor Essential Revisions

15. Results for the stratified models were mentioned throughout the text but only one was shown. Why was the TC group not included as there were only two tables?

A table with results for the TC group model was not included because only one variable remained in the model after the backwards-stepwise elimination process. Hence we believe that this data can be conveyed adequately and more succinctly in text.
**Discretionary Revisions**

16. Multivariate regression is often misused instead of multiple regression. Multivariate regression model imply that there were more than one dependent variable in the model such as MANOVA procedure.

We thank the reviewer for identifying this oversight. The three instances in text where we referred to a multivariate regression have now changed so that the correct terminology, multivariable regression, is now used (lines 32, 262 and 266).

17. The titles for Table are ambiguous, especially Table 1. It should refer to the completers used for this analysis. The Logan Health Living Program was referred as the origin of the data.

We agree, and have changed the table headings. Further elaboration within the headings was limited by the word limit for table headings (15 words).

*Table 1: Characteristics of Logan Healthy Living Program participants with complete data at baseline and follow up.*

*Table 2: Predictors of a meaningful increase in physical activity from baseline to 12-months (usual care group)*

18. Have the authors consider doing a sensitivity analysis using all the data with missing data imputed, to see if the results still hold?

In response to the Reviewer’s comment, we did conduct a sensitivity analysis imputing no change in physical activity (i.e. unsuccessful at increasing by 60 minutes per week) for participants with missing physical activity data at 12 months. The results of this analysis were not different from the original analyses for the TC group. The sensitivity analysis for the UC group yielded similar outcomes to those seen in the original analysis. Although fewer variables were included in the final model, two sociodemographic variables, education and employment, were identified as being the most important predictors. A transparent account of these analyses has been added to the manuscript (lines 341 - 352):

“The associations between baseline marital status or BMI and meaningful physical activity improvement are inconclusive; however it is possible that participants who are not married or who are overweight may respond favourably to minimal intensity interventions. Participants who are married may have greater time constraints related to home or family life, leading to an actual or perceived reduction in the amount of time available to dedicate to increasing physical activity [51], and having a higher baseline BMI has been identified as a predictor of improvement in physical activity in several studies [52]. However these findings, particularly the association between physical activity improvement and BMI, should be interpreted cautiously.”
UC group participants with a higher BMI were more likely to drop out of the trial; therefore it is possible that the association between baseline BMI and a meaningful improvement in physical activity may be an artefact of selective drop out, with BMI actually being a predictor of study retention."

Please note that in the process of conducting the sensitivity analysis, a small error was identified in the analysis. This had lead to the wrong number of UC group participants being included in the model. This error was amended and the statistical analyses were redone, with no major changes in the results being reported.