Author's response to reviews

Title: Applied public health research - falling through the cracks?

Authors:

Rebecca K Simmons (rebecca.simmons@mrc-epid.cam.ac.uk)
David Ogilvie (david.ogilvie@mrc-epid.cam.ac.uk)
Simon J Griffin (simon.griffin@mrc-epid.cam.ac.uk)
Lincoln A Sargeant (ls348@cam.ac.uk)

Version: 3 Date: 13 August 2009

Author's response to reviews: see over
Dear Dr Norton,

Re: Applied public health research – falling through the cracks?

Thank you for inviting us to revise our paper and re-submit to *BMC Public Health*. We are delighted that the reviewers think our paper presents an interesting and relevant case study, which highlights an important issue in public health research at the moment.

Please find attached a revised version of the paper with changes highlighted in red, along with a detailed response to each point raised by the reviewers below.

Please do not hesitate to contact me if you require further information. I look forward to hearing from you.

Yours sincerely,

Dr Rebecca K. Simmons  
Career Development Fellow
Referee One

Major compulsory revisions
1) Page 5 – The ethics of randomisation.
This argument is underdeveloped. At present it looks like it is based on personal opinion ‘we considered it unethical’. Many would argue (and there is a substantial body of literature on both sides of this debate) that, as there is little robust evidence that exercise referrals schemes ‘work’, there is sufficient clinical equipoise to consider randomisation ethical. There is a common line of argument amongst RCT proponents that it is only unethical to deprive the control group of a service if you are confident that service is of benefit to them – a position which has yet to be reached with exercise referral, given the current evidence base. The authors need to make clear how they would counter these arguments. Is this because the service exists already? What would be the implications of dismantling the service and setting up a randomised trial?

We have strengthened our argument about the ethics of randomisation by adding the following points to our article:

Discussion, page 7: We were...concerned with the ethics of deliberately withholding an established NHS service. Sowden and Raine (2008) concur with our findings by arguing that the experimental evaluation of exercise referral in England is now unlikely [4] because of the widespread assumption of effectiveness, the comprehensive coverage of the schemes, the indirect adverse consequences of dismantling the schemes, and the lack of appropriate process and outcome data.

Discussion, page 8: In addition, RCTs usually start from an assumption of equipoise — i.e. a position of not knowing which of two competing interventions is more effective, or not knowing whether an intervention is likely to be beneficial or harmful [17] — but in practice, previous research may suggest otherwise. In this case, a recent systematic review based on 18 studies (including six RCTs) has identified that exercise referral has a small but significant effect on increasing physical activity in some people [18]. The authors recommended that more research was needed to understand uptake and adherence, and also suggested the use of well-conducted qualitative studies. Furthermore, although physical activity has been consistently shown to be imprecisely measured by self-report [19], none of the RCTs in the systematic review incorporated objective measurement, a limitation which our proposed study design would have overcome. Methodologically weaker RCTs (for example, those with small numbers of participants, low and differential retention rates, imprecise outcome measures, and lack of attention to allocation concealment) should not necessarily “trump” methodologically stronger observational studies [20]...As such, the quality of individual studies is now receiving greater emphasis in the formulation of evidence-based guidance than was originally the case [21].

Discussion, page 9: In general, there is a need to broaden the scope of the criteria that are used to appraise and evaluate public health interventions, including the use of qualitative and observational data [4, 15, 16]. In some situations, as with our proposed evaluation of an existing service, an RCT maybe neither an ethical nor a pragmatic choice, nor does it necessarily provide the most relevant information or the most unbiased estimates of effects. Evidence-based public health must therefore rely on a variety of types of evidence, often in combination. As Victora et al/ state “...randomisation, without further analyses for adequacy and plausibility, is never sufficient to support public health decision-making, regardless of the level of statistical significance achieved” [13].
2) Page 5 - Recruitment and attrition.
I believe that most exercise referral RCTs have seen fairly equal attrition between intervention and control groups rather than large control group attrition as you anticipate here. Why do the authors anticipate that their study would be different? This statement needs further clarification and justification.

Our expectation that the control group would have suffered larger attrition rates than the intervention group is based on consultations with exercise referral staff and service users prior to the start of our evaluation. The difficulties in recruiting and maintaining a non-intervention or delayed-intervention control group have also been outlined in Isaac's EXERT trial of exercise referral (2007), a reference to which has now been added to the text. Referee Three appears to support our assertions when referring to the difficulties of using waiting list control groups in exercise referral evaluations.

3) Page 7 – Efficacy versus effectiveness.
I would be careful about suggesting that RCTs are interested only in efficacy and not in real world effectiveness. This ignores the move towards pragmatic randomised controlled trials (such as the current evaluation of the National Exercise Referrals Scheme in Wales), which attempts to embed RCT methods into the scheme’s rollout. Here and elsewhere in the article (e.g. page 5), I feel that the authors need to put a little more thought into explicating the trade-off between internal and external validity, which appears to be one of the main cruxes of their argument. In what ways would one expect a randomised controlled trial to lack external validity by comparison to a simple before and after design? Can the authors think of any examples? There are some highly relevant recent articles, such as one by Gidlow et al. (2008) which make points similar to those by the authors, but are not discussed.

We agree with the reviewer about the importance of the discussion concerning the spectrum from efficacy to effectiveness, and the apparent trade-off between internal and external validity (albeit some trials have limitations in both internal and external validity).

It is generally accepted that a study must be internally valid before we consider applying the results. However, an RCT does not guarantee internal validity. If the RCT is too small chance could still be a factor. If the blinding is not adequate for the intervention being studied bias can still be a factor. Randomisation will balance confounders only if the study is large enough. For service developers and users such as the NHS, a key concern is how to implement the results of RCTs in practice. Many RCTs have highly selective inclusion criteria, and even among the eligible population have low recruitment rates (if indeed they are even reported), leading to unrepresentative study samples. In the context of exercise referral schemes, people who agree to be randomised to the possibility of not receiving the service, or who agree to allow their GP to refer them to such a trial, are likely to be systematically different from those in which the service providers decline to permit an RCT. We have extended the Discussion section (please see page 9) to include further explanation of this issue.

Minor compulsory revision
4) Page 7 – Stakeholders dependence on continuation of service.
Some clarification here is needed of the authors relationship with the stakeholders and implications for the research process. Degree of independence between the study and the scheme. Likely pressures for positive results?
At the time the DH best practice guidance was released, targeted exercise referral schemes in Cambridgeshire were already well established and being utilized as part of the management of medical conditions. However, local NHS colleagues expressed a need to respond to NICE guidance regarding further research on the effectiveness of exercise referral. A service–academic partnership group was therefore charged with developing a rigorous evaluation study of local exercise referral schemes. This group included representatives from Primary Care Trusts, District Councils, CamSTRAD (the local support team for NHS research and development) and exercise referral schemes in Cambridgeshire, as well as academic partners from the Department of Public Health and Primary Care at the University of Cambridge and the Medical Research Council Epidemiology Unit. This group, who included the authors, met monthly to develop the proposed evaluation, and coordinate the pilot study. The group met independently of leisure centre staff, who were responsible for delivering the intervention and taking measurements at baseline and 12-weeks.

As outlined in the text, leisure centre staff were trained in anthropometric, physical activity and fitness measurement and standardised equipment and operating procedures were used to ensure valid and reliable data collection. Training was carried out by MRC Epidemiology field epidemiology staff, independently of the steering group. Furthermore, daytime physical activity was measured objectively using an accelerometer, and venous blood samples were collected by local GPs, in order to further reduce potential bias.

Given these points, we believe that there was a sufficient degree of objectivity and independence between the authors, the leisure centre staff members taking measurements, and study participants. In the interests of keeping the article succinct, the bulk of this information has not been included. However, if the editor indicates that it is necessary, we would be happy to add further text to this effect in the paper.

5) Page 7 – Dissonance between types of evaluation required and types research bodies. This is mentioned several times throughout the manuscript. However, it could be made clearer what it is that the authors mean by this statement.

We introduce the idea of dissonance between the types of evaluative research required by commissioners, organisations developing evidence-based guidance, and those which research funding bodies are prepared to support in the Background. This point is discussed and developed in some depth throughout the Discussion (from the second paragraph onwards). Given the increasing length of this article, we would prefer not to expand on this point further if the editor is in agreement.

Discretionary revision
6) Page 6 – Non-randomised study (or quasi-experimental design).
Some good points here. In addition, what incentive would patients in control areas have to enter the study if they knew at baseline they had no chance of entering the service straight away or (if put in a wait list control group) after a 12 month waiting period?

We thank the reviewer for this suggestion and have added the following sentence to the relevant paragraph on Page 6: “In addition, there would have been little incentive for individuals to enter the control arm of such a study if they had no chance of receiving the service or if they were to be added to a long waiting list.”

Referee Two

Discretionary Revisions;
Additional insight into the stakeholder viewpoint would be useful and highlight the tensions faced by those working in partnership to evaluate public health programs.

We thank the reviewer for their helpful suggestion. We recognize that stakeholder viewpoints are an important component of public health evaluations, and hope to have highlighted the major tensions we faced in the development of our own evaluation (see pages 5, 6 and 7). Given the increasing length of this article, we would prefer not to expand on this point if the editor is in agreement.

Referee Three

It would be helpful for the reader to have a sense of the equivocal findings of the systematic review used by the PHIAC to derive their recommendations. These 4 studies were assessed for their quality and summary findings. The original reviews are available from the NICE website. It would also help to have a brief overview of the actual process used for the construction of these particular recommendations. The findings of a systematic review were assessed by a committee of general public health specialists (none with specific physical activity expertise). The committee discussed their conclusions related to the recommendation, asked additional questions of the reviewers and considered the economic report as well. It is too simplistic to blame NICE for the conclusions of an independent committee and it may be a reflection of that committee’s opinion at that time. Perhaps when this review is updated (due 2009-2010) their recommendation may be overturned.

We thank the reviewer for their suggestion. Rather than comment specifically on the equivocal findings of the systematic review used by the PHIAC to derive their recommendations, we have included some text to address the issue in more general terms. We hope that the following paragraph (page 10) makes it clear that we not blaming NICE for the conclusions of an independent committee, but rather, that we acknowledge the difficulties in producing guidance on public health topics where there is limited evidence:

“Producers of guidance, and commissioners and practitioners receiving guidance should have due regard for the quality and quantity of the evidence underpinning recommendations before acting on them, since guidance development groups working on public health topics are often in the difficult position of having to make recommendations based on limited evidence. It may also be helpful for bodies issuing evidence-based guidance, such as NICE, to specify acceptable study designs in their recommendations for further research, make clear the conditions under which interventions should or should not be implemented, and explain what health care providers should do while the research called for is being completed. In this context, where a recommendation is made not to offer an intervention outside the setting of a controlled research study because evidence for its effectiveness is described as ‘equivocal’, it should be made clear whether the equivocation is between a beneficial effect and a harmful effect (in which case the condition of equipoise may be satisfied) or between a beneficial effect and no effect (in which case the condition of equipoise may not be satisfied and commissioners may reasonably decide, in the absence of a more effective alternative course of action, to give existing services the benefit of the doubt until the evidence base is populated with stronger evidence either way). Costs should also be considered when describing results as equivocal. If a service or the evaluation of a service is extremely expensive, this may also affect the condition of equipoise.”

It would be useful to discuss the design and results of the Issacs et al HTA exercise referral trial in light of the authors’ views that the design of a RCT with a wait list is unethical or unpopular. The difficulties this trial had in recruitment supports this view.
Thanks to the reviewer for this reference. We have added a sentence to the bottom of page 5 and cited the Issacs reference in order to support our view that a waiting list control group would have been difficult to recruit and maintain. Please see also our response to Point 2 from Referee One.

Minor Essential Revisions
Reference 5 has been updated in 2008 with the same conclusion.

This reference has been updated.

Reference 8 is not related to the NICE recommendation process please amend.

This reference has been omitted.