Author's response to reviews

Title: Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions.

Authors:

Colin J Greaves (Colin.Greaves@pms.ac.uk)
Kate E Sheppard (kate.sheppard@pms.ac.uk)
Charles Abraham (s.c.s.abraham@sussex.ac.uk)
Wendy Hardeman (wh207@medschl.cam.ac.uk)
Michael Roden (Michael.Roden@ddz.uni-duesseldorf.de)
Philip H Evans (philip.evans@pms.ac.uk)
Peter Schwarz (Peter.Schwarz@uniklinikum-dresden.de)

. The IMAGE Study Group (Peter.Schwarz@uniklinikum-dresden.de)

Version: 2 Date: 22 October 2010

Author's response to reviews: see over
Re: MS: 6076024914240378 “Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions” (revised title)

22\textsuperscript{nd} October 2010

Dear Sir /Madame,

Thank you for invoking us to re-submit the above manuscript. Please thank your reviewers for their considered and helpful comments. We have submitted a revised copy of our manuscript with Tracked Changes via your online submission service and the point by point response to the reviewers comments follows.

We look forward to receiving your decision.

Yours sincerely,

Corresponding author:

Colin Greaves PhD C Psychol
Senior Research Fellow
Response to reviewers’ comments:

Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions.

Firstly, we would like to thank the reviewers for their detailed and helpful comments. We feel that the article has improved considerably as a results of addressing the issues raised.

Reviewer 1:

We were pleased that the reviewer found that our article addresses an important question, that our review was carefully conducted and comprehensive, and provides a valuable contribution to the literature.

Specific comments:

Methods:
1. Selection criteria (p.4):
   a. Clarify difference between systematic review of individual level interventions and systematic reviews of e.g. RCTs, which can be individual level interventions, too.

   The inclusion criteria specify the type of intervention of interest (individual level interventions) and type of study (RCTs, quasi-experimental studies etc). This includes studies that are RCTs of individual level interventions. We have edited the presentation of the inclusion criteria to make this clear and also to clarify what we mean by ‘individual level interventions’ (i.e. interventions delivered to individuals either singly or in group sessions, but not whole-community or whole-population level interventions such as media campaigns or changes in the local environment). It is hard to make this distinction without using the word ‘individual’ which may then be confounded with the notion of one-to-one (as opposed to group-based) intervention. If preferred, an alternative would be to remove the term “individual-level interventions” and replace it with something like “individual or group based interventions”.

   b. Define “high cardiovascular disease risk score”. Which scores and thresholds were used?

   We specified this in the protocol (any validated cardiovascular risk scoring system, such as QRISK or Framingham would have been acceptable), but in practice, no reviews identified this as a target population. However, we have provided an example in the text to clarify our intention. NB: this criterion was included because of the strong relationship between cardiovascular and diabetes risk factors.

2. Data extraction (p.5): clearly define the content of these pre-defined intervention categories within the methods section.
It is quite hard to do this as the definitions /categorisations used could (and did) vary between the reviews included. We have tried to provide a brief ‘meta-definition’ in each case, but within the reviews we extracted details of any analyses that related to theoretical basis, behaviour change techniques, mode of delivery, provider, intensity, target population and setting. As reported in the Discussion, variation in conceptual definitions used in the literature (e.g. of “established behaviour change techniques”) is a problem in the literature.

3. Outcomes (pp. 4-5): Authors define primary and secondary outcomes. However, as at least 4 primary outcomes are stated (weight in kg or BMI, met-hrs per week, and frequency) not considering the different follow-up durations, this distinction seems somewhat problematic.

This was not reported very well and we have clarified the difference between the aims of our review and the aims of the reviews we examined. Individual reviews which we included may have had primary or secondary aims (e.g. to establish the effectiveness of a particular type of intervention on a particular outcome). However, our aim was much more exploratory – to summarise the evidence relating the content of interventions to their effectiveness in producing behaviour change (as evidenced by either change in weight or direct measures of behaviour). To achieve this, we examined reviews which had defined weight loss or behaviour change (dietary or physical activity behaviour) as a primary outcome. As we did not meta-analyse this data, but sought only to summarise (and grade) the evidence, the idea of primary /secondary analyses does not really apply to our own synthesis. We have therefore amended the text here to clarify that “We selected reviews where the primary outcome measure for dietary interventions was weight or weight loss (kg or Body Mass Index (BMI)) or physical activity.”

We agree that the time-point at which outcomes are assessed is an important issue and we have accounted for this in summarising the data (“.. at a median 6 months of follow up” etc).

4. Although this study is a systematic review of reviews, I believe a PRISMA checklist should still be used and provided.

This is a reasonable request and we have sought to do this (see new Supplementary document on this)

5. Clarify abbreviations: e.g. BMI, met-hrs; once abbreviations introduced use consistently e.g. RCT instead of randomised controlled trial

We have sought to address this throughout

Results

6. Search results (p.6):

a. Please provide more information regarding the results of expert contacts, how were experts selected, how many contacted, responded etc.?

This process was not formally recorded. We selected experts based on existing contacts of the co-authors and by approaching authors of key reviews and guidelines. We also invited the 70+ members of the IMAGE Study Group to identify
possible papers. We particularly approached contacts in the area of health psychology to identify papers relating behaviour change techniques to outcomes. Approaches were made both in conversation and by email and by presentation of our planned review at meetings (e.g. PSAD Europe, IMAGE Study Group). We did not record how many experts were approached or how many responded, although it is fair to say that the overall response rate was quite low. However, this process did identify 2 reviews that were unpublished at the time and where the authors were willing to share their data with us (one is now published and one in press).

b. More details regarding the number of excluded studies due to insufficient quality or other exclusion criteria would be interesting. This differentiation could be included in the flow-chart.

We have included a table of excluded papers in the supplementary materials (Table S14) and we have added details to the flow chart.

7. Supplementary tables:
   a. Abbreviations should be explained beneath tables.
   
   We have sought to address this in the revised tables.

b. S6 -S13 seem to contain some inconsistencies with regard to the way, effect measures/ effect sizes are presented. In addition to the already existing differences between included studies, this causes even more confusion.

We have sought to address the consistency of presentation of figures as much as possible (e.g. format for reporting of 95%CIs, SMDs etc and clarifying where we are referring to effect size in terms of a standardised mean difference as opposed to reported changes in outcome measures). We have reported the results (SMD or change in outcome measure) that were reported in each review, rather than seeking to standardise the type of effect reported (SMD or outcome or both) across all the tables. This is because the exact form of analysis used to derive the result was not always reported clearly in the original study and we might introduce errors by trying to standardise the format of the results.

8. Evidence synthesis (pp. 7):
In general, the amount of presented data is substantial and the current presentation somewhat difficult to follow.

We have attempted to introduce more structure in the reporting of the Results by ordering the evidence presented under sub-headings in both the main text and the evidence tables. In some cases this was not possible as some reviews conduct analyses which span sub-headings (e.g. in the Population Characteristics table), but we feel that the changes made have introduced a much clearer structure overall.

a. Presented outcomes seem to differ from definitions in the methods section. E.g. overall effectiveness: authors initially present weight loss, followed by reduction in diabetes incidence and physical activity behaviour. Diabetes incidence was defined as a secondary outcome and should be presented in this way, if authors decide to
keep this distinction. Within the following findings of intervention components secondary outcomes were not further consistently reported. Instead the duration of follow-up, which was not defined as an outcome

The reference to primary/secondary outcomes has already been clarified (see comment 3 above).

In the revised Methods section we have presented the main outcomes of interest as: Weight, Physical activity, Diet and Other outcomes of interest. Accordingly, the outcomes presented in the overall effectiveness table and the Overall Effectiveness section of the results have been re-structured under headings of Weight Loss, Physical Activity, Dietary Intake and Other Outcomes. This re-structuring should make the description easier to follow and to relate to the evidence in the Tables.

b. With regard to the duration of intervention effectiveness I was wondering whether this could be presented in terms of e.g. short term (e.g. up to 12 months) and long-term (more than 12 months)?

It is impossible to break down the findings in many different ways without losing coherence. We have indicated the time period for the outcomes reported in all cases and there is also an important statement about changes in outcomes over time (now under the ‘Other Outcomes’ heading).

c. Theoretical basis (p. 8): evidence of effectiveness is being stated without providing information regarding the outcome(s) in question.

This has been corrected.

Discussion

9. Effectiveness (pp. 13-14). Currently, the magnitude of intervention effects is almost exclusively reported in supplementary tables. In addition to the evidence of an effect, a presentation of a range of the magnitude of intervention effects (e.g. in terms of proportion being active, weight loss) would be very useful.

We have introduced more text on the magnitude of effects in the Results section and in the text relating to overall effectiveness in the Discussion (para 2).

10. Please discuss whether findings and conclusions find confirmation in the light of evidence from other target populations (not at risk of diabetes), but e.g. healthy adults, people with diabetes, cardiovascular disease etc. Can these conclusions be applied to other target populations?

We have added a paragraph to the end of the first section of the Discussion on this.

11. Regarding the frequency of interventions/contacts (p. 14 first paragraph) it would be useful to state a number or frequency (or a range) of contacts in the conclusions/discussion. A plain statement of more is better does not help much.

Unfortunately the evidence is not precise enough to suggest any specific threshold or cut off point. However, to provide some kind of anchor point, typical ranges of
contact time, number of sessions and intervention duration have been provided in the Intensity section of the Results.

12. Strengths and limitations (p. 14): Identified systematic reviews likely include, at least in part, the same primary studies. This might result in double counting or overrepresentations of certain findings. How often was this the case? Please discuss implications.

We have added some text in Strengths and Limitations with regard to the interpretation of analyses which are in some cases based on overlapping sets of studies. Importantly, the multiple analyses conducted are in the main mutually supportive (albeit with different levels of rigour/quality around the reviewing and analytic methods used). Areas of disagreement between reviews have been highlighted in the text (as with the issue of intensity where clearly more evidence is needed). We have suggested (through the way we have graded and presented the data and in Strengths and Limitations at the end of para 2) that the key to interpretation of this wide-ranging evidence base is to consider the highest grade of evidence available for any particular statement rather than the number of separate analyses which support it.

References
13. WHO reference (2): access date?
14. Reference 13, 18, 29, 35: check format

These references have been reformatted as appropriate and the remaining references have been checked to be compliant with the journal’s preferred style.
Reviewer 2:

*We were pleased that the reviewer felt that our review covers a very important topic, and is clearly written and comprehensive.*

Minor essential revisions required.

Nil

Discretionary revisions as follows

1. Data extraction paragraph 1 You may wish too clarify how you chose the pre-defined intervention components

*We could add that “These were selected based on consultation within the IMAGE Study Group and knowledge of prior reviews and guidelines on weight loss, physical activity and dietary change”. However, we are not sure saying this would add much value to the paper.*

2. Results/ Behavior change techniques / paragraph 3
You begin with ‘evidence indicated that change in diet and physical activity was greater for a) interventions which used established behavior change techniques etc....’ and you provide a weight loss and physical activity outcome (additional weight loss of 4.5kg at 6 months) etc.. It is not clear what this weight loss relates to and is in addition to. Is this compared to controls receiving no advice or controls receiving non established behavior change techniques?

*This is based on experimental manipulation of the use (vs. non use) of established behaviour change strategies in addition to other techniques such as providing dietary advice. This is causal evidence, which demonstrates that adding behaviour change techniques (e.g. to the simple provision of dietary advice) increases weight loss by 4.5Kg at a median 6 months of follow up. We have edited the text to clarify this.*

3. Results /Intensity/ paragraph 1 It would be important to have stated in the manuscript how many good quality low intensity interventions were included in analyses. My understanding is there would be very few high quality, low intensity interventions to compare with and therefore I wonder if the ‘favoring of high intensity interventions ‘ might be due to lack of available comparisons. For example are any internet based interventions included or available in previous reviews? The reference from Shaw indicates there was heterogeneity in the studies and the contact ranged from weekly to monthly with a median of weekly which is very intensive contact. I think your conclusions are probably correct, but as you are focusing on a target group at risk of developing diabetes I think it important not to discount low intensity interventions as a feasible option unless it is clear they are not helpful rather than just not available. It would be important to discuss this more thoroughly, particularly in conjunction with a discussion around the other components of intensity ( other than contact) you analyzed.
It is an interesting hypothesis (that the association between low intensity and lower effectiveness might be confounded by lower quality in the low intensity interventions. However, we are not able to comment on the quality of individual interventions here as this is a review of reviews, not of individual studies. We did not therefore rate the quality of individual studies (although we did give higher quality scores to reviews that took study quality into account in their analyses). A comment on this is now provided in the para 2 of the Discussion. We have also added a comment on the need to unpick the different elements of intensity in the discussion section on Directions for Future Research. Against this hypothesis however, there is the causal evidence (1+) from Shaw et al (2005) which meta-analyses 10 trials comparing low and high intensity versions of the same intervention. However, the definition of intensity in this study was quite odd (including number of behaviour change techniques) and so we have not put too much weight on this analysis.

We feel that, by presenting the findings on intervention intensity as an ‘association’ rather than as causal evidence and in presenting the Recommendation relating to this as grade B (in Table 2), we have been suitably cautious.

We agree that the evidence does not preclude the possibility of an effective low intensity intervention, such as internet-based or other self-administered approaches. However, we did not find any sufficiently high quality reviews of internet-based interventions and this gap in the literature (and the need for more effective lower intensity interventions) is now highlighted in the discussion section under Strengths and Limitations and under Directions for Future Research.

4. Discussion Paragraph 3 ‘we found interventions can be delivered successfully by a wide range of providers in a wide range of settings ….. and can be effective for a wide range of ethnic and age groups.’ However your results suggest there is little evidence in ethnic groups, stated as ‘there was very little review led evidence on the relationship between ethnicity and intervention effectiveness.’ You may wish to clarify this statement or alter your conclusions.

We have attempted to clarify this in the text. Although no clear associations were found between ethnicity and effectiveness, there were clear examples of interventions that were effective in a number of ethnic groups (see evidence tables). This has also been referred to as an area where further research is needed.

5. Conclusions. Point e) As I have stated above I would like further evidence before recommending higher frequency or total number of contacts. Possibly the evidence suggests ‘frequency is important, along with duration and more behavior change techniques’.

We agree that the intensity recommendation should be less definitive than the others, and have re-stated the Conclusion accordingly to say that interventions may also benefit from providing a higher frequency or total number of contacts.”

Reviewer 3

We were pleased that the reviewer commended us on a coherent search
strategy and a thorough quality assessment of the reviews.

Major Compulsory Revisions
Key points:

1. The authors need to make it absolutely clear in the abstract and title that this is a review of reviews, at present it reads as though the authors have conducted 129 analyses relating intervention components to effectiveness. Please make it clear that this is purely a summary of the analyses of the included reviews, so as not to mislead potential readers.

The title of the paper has been amended to reflect the fact that this paper is a review of reviews: The abstract has been amended to include clearer reference to the fact that this is a review of reviews and that it summarises and grades analyses from other reviews.

2. Overall the authors should be more tentative in their interpretation of their findings and the subsequent recommendations offered. The review is extremely comprehensive, covering a number of different types of systematic reviews covering different populations, behaviours and analysing a variety of outcomes. As such there is a large amount of heterogeneity between the studies included, thus recommendations given on the basis of comparing and summarising these should be tentative. Particularly in the case of where the argument is backed up solely by the results of one systematic review.

It is the authors’ view that our conclusions are appropriately worded to reflect a) their causal or associative nature and b) the quality (not quantity) of the underlying evidence. We have graded the evidence using strict quality criteria and sorted it according to causal evidence (where the issue at question has been experimentally manipulated) and associative evidence (where the evidence is based on associative analyses such as meta-regression or making comparisons across, rather than within, studies). There was in fact very little heterogeneity between the reviews studied in terms of the patterns of data presented. Where there is disagreement or lack of evidence we have highlighted this.

However, we agree that in the conclusion we should have been more cautious about the issue of intervention intensity (see also our response to Reviewer 2, Comment 5) which is based on lower quality causal evidence and associative evidence. This has been re-worded accordingly.

3. The primary and secondary outcomes of this review are not clearly defined, I would suggest these are made much clearer to aid comprehension. Specifically, the authors state that the secondary outcomes are for both types of intervention; dietary and physical activity. However, self-reported change in dietary behaviour surely should be a secondary outcome of dietary interventions only, and likewise cardio-respiratory fitness should be an outcome for PA interventions only?

Please see our response to Reviewer 1, point 3 above. We have amended the text here to clarify that “We selected reviews where the primary outcome measure for
dietary interventions was weight or weight loss (kg or Body Mass Index (BMI)) or physical activity.

4. The authors do not differentiate between evidence of absence and absence of evidence. Specifically it is unclear if there is evidence which suggests there is no association between a particular characteristic i.e. theoretical basis and outcome, or if there is presently no evidence assessing whether such an association exists. This differentiation should be made clear throughout the manuscript.

Where there is clearly a lack of any evidence, we have included comments saying “.... there is a lack of evidence for....” or “....there is a lack of good quality evidence...” (e.g. Ethnicity section, Setting section in Results). However, this is not always a categorical or easy distinction to make. In some cases there is a lot of evidence implying no effect, but none of it is causal or of high quality (e.g. for differences in intervention provider), so it would be hard to conclude definitively that no such effect exists. In no case have we suggested that there is definitive evidence of no effect.

In all cases, we have presented an overview of what evidence is available and at what level of quality, so readers should be able to judge for themselves where there is evidence of absence or an absence of evidence. We have also highlighted in the Discussion where, in our view, further evidence is needed.

Introduction
5) Please describe in what way current interventions are too intensive, and where is the evidence? Further, what diabetes prevention programmes are out there, are they currently ineffective or just too intensive?

Text on this issue has been added to the Introduction.

6) I think it could be clearer here about what is referred to by intervention components and characteristics, does this mean intervention techniques e.g. goal setting, or other intervention characteristics such as setting, delivery mode etc. Or is it both?

Both: We have tried to add greater clarify to this by changing the text in the Background to “In developing such programmes it is critical to ensure that the intervention components (i.e. behaviour change techniques and strategies) and characteristics (e.g. setting, delivery mode, intervention provider) most strongly associated with effectiveness are retained.” and also by giving examples of the components of interest (see revised section on Data Extraction).

7) Is there any reason why the review is focusing on individual level interventions only? Is there evidence that these are more effective/cost-effective?

We have clarified this point by removing the phrase ‘individual level intervention’ in the background and providing a clear definition in the Type of study section in Methods. We are using the phrase ‘individual level’ to distinguish between interventions delivered at the community or societal level (e.g. changes to transport infrastructure or the environment or media campaigns) and interventions delivered to individuals (this can be face-to-face in groups or one-to-one, or in self-delivered
modes such as via the internet or self-help books). This does not refer to the contrast of individual based vs. group based mode of delivery. See also our response to Reviewer 1, Comment 1a above.

Methods
8) Can the authors explain why the search was only for the years 1998-2008, were there no reviews in this area pre-1998 or was this for pragmatic reasons?

The selection of the literature search timeframe was for pragmatic purposes. The reviews we selected had searches dating from 1966 through to 2008 (this is now stated in Review Characteristics in the Results section), so the strategy encompassed a broad spectrum of studies. This detail can also be seen in table S1 - Characteristics of Included Reviews.

9) The PICO Criteria needs to be made substantially clearer to make the inclusion and exclusion criteria more explicit. Further, the review specifically mentions cardiovascular disease twice in the inclusion criteria but there is no mention of this disease in the introduction, what relevance does this have to type 2 diabetes interventions? The link between type 2 diabetes and CVD needs to be explicit for readers who are not experts in this field.

The risk factors for type 2 diabetes and cardiovascular disease are largely shared. We have dealt with this by removing the text “and/or cardiovascular risk” from the Study Populations section in Methods.

10) A little more detail on the use of the SIGN grading system would be appropriate here, although I am aware there is much more detail included in the supplementary material Table S5. Many readers may be unfamiliar with this grading system, and given the weight it has on the final study findings, it is essential this is carefully articulated in the methods section. Specifically the sentence starting ‘quality of the evidence for each analysis is described as……’ should be reworked. Please state what the ++, + and – refer to and why, and which types of review fall under which category (causal/associative).

We have added more detail on SIGN grading into the paper in the Grading of evidence section (Methods) and have tried to rephrase the sentence in question to add clarity.

Please note that, rather than the ‘type of review’, it is the type of analysis that is graded. This means that within the same review, different analyses may be given different grades.

11) I am unsure whether the title of Evidence Grade 1 as ‘causal evidence’ is correct. According to the authors’ description, reviews would be categorised as this if they included RCTs and which statistically compared two randomised groups either descriptively or by meta-analysis. You cannot assume that the studies within these meta-analyses conducted mediation analyses for the particular characteristic under consideration, which therefore makes it difficult to infer a causal link between characteristics and effectiveness as a consequence of a review only. Consider
changing this to reflect what actually differs between the two gradings (1 & 2), which from my understanding refers to the type of review included.

_Evidence was only graded as level 1 if it was clear that the analysis in question synthesised data from RCTs that experimentally manipulated the component or characteristic of interest (e.g. trials that randomised people to receive the same intervention either with or without a social support element). Where we have claimed the evidence is causal, it definitely is._

Results:
Overall the results section could be substantially reworked to aid the readers’ comprehension, in particular:

12) The authors may want to think about removing the low quality evidence from this section as I am unsure this really adds much weight to their argument. Particularly as on a number of occasions it is stated that the review is focusing on higher quality evidence.

_It is a reasonable observation that the ‘very low quality’ evidence in particular does not add much, particularly where there is other evidence that is stronger. Highlighting the areas where only low quality evidence exists can inform recommendations for future research. However, in all cases here the very low quality evidence only exists alongside higher quality evidence, so we have taken the reviewer’s advice and removed it. We have added a comment to the Grading of evidence section (Methods) stating that this evidence is not presented._

13) Where possible please include effect sizes and/or odds ratios, particularly of significant findings in the results section. Readers may not have time to look through all the tables for this information.

_We have added data indicating the size of effects where possible and where this might help interpretation (particularly for causal analyses). We feel that total reporting of all the data in the tables would be a step too far however and would detract from the readability of the results. See also response to Reviewer 1, Comment 9 on this topic._

14) I am unclear on how the authors defined intervention techniques across reviews, and how they sought to evaluate the effectiveness of the specific techniques. Further, it is confusing that the authors summarise evidence of the use of ‘well defined techniques’ based on reviews included in this study and state they also used the Abraham & Michie taxonomy, why use both? There is surely some overlap in the techniques; self-monitoring, relapse prevention, barrier identification identified in your the reviews are certainly also included in the taxonomy.

_We did not define intervention techniques across reviews, we have simply reported what definitions other reviewers have used and attempted to summarise the analyses that were presented based on their definitions. We did not use the cited taxonomy to code data or to inform our review per se. We have provided this reference so that readers can look up definitions of the specific techniques mentioned. We have altered the text on this (Behaviour Change section of Results) to try to make this situation clearer._
15) I am unsure why the Michie et al (2009) meta-regression of effective techniques in healthy eating and physical activity interventions (ref. 31) has not been included in the section on behaviour change techniques. I am unsure why this has not been included here as this paper provides evidence of specific behaviour change techniques associated with effectiveness. I would suggest this is included in these analyses.

The evidence in this review features several times in the Theoretical Basis and Behaviour Change Techniques sections of the Results and the Supplementary Tables.

16) On page 9 evidence from the Ogilvie (2007) meta-analysis is referred to as medium quality, yet on page 8 evidence from the same meta-analysis is described as low quality. Can the authors explain this?

We graded individual analyses (not reviews) so therefore it is perfectly possible for two analyses from the same paper to have a different grade. It will have depended on the type of analysis, the way it was conducted (methodological issues) and the nature of the evidence involved. In this case, we were looking at 1) a descriptive summary of 27 RCTs and observational studies (with data for the 17 RCTs reported separately) of walking intervention versus control (Table S6: grade 1- evidence due to heterogeneous interventions and population inclusion criteria and reliance on mainly self-report data); and 2) a descriptive summary of 6 RCTs of pedometer based interventions versus control (Table S8: grade 1+, tighter inclusion criteria /generalisability, mostly using objective measures of steps /day (e.g. from blinded pedometers used alongside the intervention devices). Hence, both these analyses were causal (grade 1), but the evidence in one case incorporates more risk of bias than the other (hence – as opposed to +).

17) Although the authors state that interventions that included certain techniques were effective/ ineffective at bringing about the desired change, they are unable to say that it was the particular technique itself that was effective/ ineffective unless moderator analyses were conducted in the reviews. Other intervention characteristics might have moderated the impact of the behaviour change technique itself. Please make sure this is clear both in the results and the limitations section of the discussion.

This is correct where analyses were associative (e.g. based on meta-regression) rather than causal (i.e. based on randomisation of participants to receive the technique of interest or not). The possibility of moderation by other covariates is the main reason why associative evidence gets a grade 2. The evidence grading takes this into account – if we say it is causal it is based on randomisation and experimental manipulation of the factor of interest. If we say it is associative, then it is based on associative analyses and is treated with suitable caution in the Conclusions.
18) In the “behaviour change technique” section of the results the authors state that change in diet and/or physical activity was greater in interventions which targeted physical activity and diet. This does not make sense.

The aim of this sentence is to highlight that changes in diet and/or physical activity were greater in interventions that targeted both physical activity and diet. We have amended the sentence to clarify this and the whole section has been re-structured to increase clarity.

19) I am unsure why the authors have described the difference in effectiveness of group versus individual interventions. It is quite clearly stated earlier in the manuscript, and in paragraph 1 of the discussion, that the review is summarising individual-level interventions.

Please see comments above (Reviewer 3, Comment 7) on this. We have tried to be clearer about this in the revised text.

Discussion
20) Para 2, line 2. Please include references to the statements about effectiveness, so it is clear which study/studies are being referred to.

These references are provided within the Results section and the evidence tables it would add a lot of clutter to repeat this for all the points summarised in the Discussion.

21) No explanation is offered for the findings, specifically in relation to previous diabetes prevention programmes mentioned in your introduction. What has been discovered that is different to what has already been found? Given that the review aimed to enhance these, how and why would the recommendations do this?

This review goes substantially beyond the existing evidence from diabetes prevention trials in terms of providing specific recommendations on the content of interventions that is associated with increased effectiveness. The existing diabetes prevention trials are focused on reporting diabetes outcomes rather than the mediating effects of different intervention components and characteristics. The meaning of the findings in relation to diabetes prevention programmes is stated in the section in the Discussion on “Implications for practice and policy”, with the text slightly revised to emphasise the relevance of the current review’s recommendations to developing more cost-effective diabetes prevention programmes.

22) A number of current diabetes prevention programmes are referenced, why are these not referenced in the introduction as this seems to be a focus of the review?

The text on these programmes has now been moved to the Introduction

23) Whilst I do acknowledge that increasing physical activity, reducing weight and improving diet can prevent type 2 diabetes, and can therefore see the logic behind the inclusion criteria; would it not be appropriate to conduct a systematic review of these interventions with moderator analyses to evaluate the effectiveness of specific
intervention components and characteristics? Instead of attempting to summarise a large heterogeneous sample of reviews of physical activity and diet interventions which are not specifically aimed at preventing diabetes.

This seems a good idea (and we did consider it at the outset), but this was not strategy we selected for this review. Our choice of methodology reflected a need to collate the evidence base quickly, to include a wide range of evidence that might be relevant and within the resources available to inform the IMAGE guideline. If we had the resources, it would be interesting to conduct an individual study level review of this literature with both meta-analyses of trials which produce causal evidence of relevance and mediator analyses (meta-regressions) to look at associative evidence. However, there would be a problem in focusing only on the diabetes prevention literature in that a) many interventions are poorly described and b) given the low number of such studies, any mediation analyses would be seriously underpowered and it would be hard to arrive at robust conclusions. For this reason we would not restrict the field to people at high risk of diabetes only. The challenge of achieving change in weight/diet and physical activity is almost identical for people with obesity as for people at high risk of diabetes for instance, so it would be a missed opportunity not to try to learn from the much larger evidence base available for people with modifiable risk factors (high weight, sedentary behaviour) and a wider range of risk factors.

24) The limitations section could be further expanded, it does not currently acknowledge the limitations of the present review, instead focusing on the limitations of the included studies.

We have included more detail about the limitations of conducting a review of reviews:

Discretionary Revisions
25) Abstract: Line 2. It is critical that the intervention……….are retained. Does this mean retained from current diabetes prevention programmes?

To clarify, we have changed the text to “included” rather than “retained”.

26) Introduction: It would be helpful to make the aims clearer. Specifically, is the aim to a) summarise the evidence of existing type 2 diabetes prevention programmes, to extract the most important intervention components from these to improve upon their current effectiveness? Or b) to summarise the evidence of physical activity and dietary interventions, not necessarily specific diabetes prevention programmes, with the expectation that effective components could be incorporated into future diabetes prevention programmes?

Option b, although this is not very well stated above a) in terms of PICO and b) we were looking for relationships between components and effectiveness not just effectiveness. We have made a specific statement of our aim in the Introduction (last 2 sentences):
27) Methods: It is worth stating who completed the searches of the databases, as has been done for the other aspects of the methods section.

This has been amended.

28) Methods: In the inclusion criteria the authors state that this review includes interventions for individuals at risk of type 2 diabetes. However in table S1 they do not detail what these risk factors are in a number of studies, indeed from the table it seems one review includes RCTs with ‘health community dwelling individuals’. How would this fit in with the inclusion criteria?

The study in question describes the included population as “healthy community dwelling individuals aged 18yrs+” so that is what we originally reported. However, the review included 13 trials of people with cardiovascular disease risk factors (e.g. overweight, high blood pressure, high cholesterol) – they were ‘healthy’ in the sense of not having any current diagnosis of cardiovascular disease, but were clearly still at risk. We have added some text to this entry in the table to clarify this.

29) Results: The authors could include a little more detail regarding the characteristics of included reviews i.e. population and delivery settings, and describe the number of descriptive, meta-analytic and meta-regression analyses.

In the section on Review Characteristics (Results) we have added some text to indicate the range of dates of RCTs included in the reviews studied (1966 to 2008). And the range of populations included in the component reviews (people with overweight, obesity, hypertension, impaired glucose regulation, etc) and delivery settings. The details for individual reviews are presented in the tables so anyone who is especially interested in this aspect can look them up.

30) Results: The authors may want to use subheadings to aid the readers’ comprehension, this is particularly in the case of “characteristics of the target population” in which gender, age and ethnicity and risk are all attended to.

This has been done. We have now organised the data under subheadings in most sections of the Results (see also our response to Reviewer 1, Comment 8).

31) Discussion: The authors state that the interventions produced clinically meaningful changes, please define “clinically meaningful”.

This is quite a complex question and we could justify this in detail with reference to epidemiological studies and intervention studies (e.g. in the Finnish diabetes prevention study, each Kg of weight lost resulted in a 16% reduction in diabetes incidence; an increase of 60 mins per week of moderate physical activity is associated with a decrease in cardiovascular risk of around 15%). However, we would prefer to keep this summary section pithy and to the point, so this is probably not the best place to provide such a justification.

Minor issues not for publication
• Asterisks of included studies are missing in the reference list.
Asterisks have now been inserted