PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A cluster-randomised controlled trial to assess the effectiveness and cost-effectiveness of an obesity prevention programme for Chinese primary school-aged children: the CHIRPY DRAGON study protocol
AUTHORS	Li, Bai; Liu, Wei Jia; Adab, Peymane; Pallan, Miranda; Hemming, Karla; Frew, Emma; Lin, Rong; Martin, James; Liu, Wei; Cheng, Kar Keung

VERSION 1 – REVIEW

REVIEWER	Dr Priya Parmar Auckland University of Technology New Zealand
REVIEW RETURNED	26-Jul-2017

GENERAL COMMENTS	This paper was nicely written and a comprehensive piece of work. Some comments are listed here: - The anticipated statistical analysis was very sparsely described in the abstract. An additional sentence to elaborate would have been preferred. - 10/22 studies which showed no, or low drop-out rates which was used to help justify their sample size calculation. The authors should have reviewed the most relevant papers from those 22 studies to identify the optimum drop-out rate or use all 22 studies to inform the expected drop out rate for this study. The authors lacked justification for using a drop-out rate of 10% based on the 10/22 studies which showed low, or no drop-outs. - Whilst the authors state they sought advice from a medical statistician and local education authorities to identify important factors in which to stratify the schools allocation to intervention or control groups it was unclear why other important factors were not used (i.e. that are associated with obesity e.g. region of Guangzhou [is diet the same or different across this city? Are all schools located within walking distance to the homes of the children attending them? What existing mechanisms are already in place in schools - do some already adhere to some physically active programme? Or have more sports-related activities? Are their extra-curricular activities such as sports available at all schools? Are some of the schools located in more remote, or rural regions of the city? Is socioeconomic status different across this city (e.g. are parents more likely to drive their children in some areas, whereas in other areas they tend to walk), are there other considerations such as the local environment in proximity of the school (are they close to parks

| | or places were physical or sporting events can take place or are they close to fast food outlets, stores which sell unhealthy foods and availability of gaming centers etc). A few are listed as potentially being recorded as potential moderators however, there is a lack of information and detail here. The authors and researchers here could have possibly made the design slightly more complex by applying this as a multi-level cluster randomised control trial.

- Were birth anthropomorphic measures collected (esp. birth weight which is predictive of later life health and well-being)

- A more detailed list of potential confounders that would be adjusted for in the statistical analysis would have been ideal

- There were no explicit inclusion and exclusion criteria. How could potential sources of contamination be minimised? E.g. siblings, relatives, friends who attend schools in a different treatment group. Is it possible the intervention effects could be diluted?

- Whilst the school was randomised per cluster, could the detail recorded per student be used for longitudinal analyses (i.e. linked with follow-up data) or will it only be summarised at the school or treatment level?

- Spelling error (page 7, line 15, should be CHIRPY DRAGON not CHIRPYD RAGON) and (page 7, line 29 'summaries' should be 'summarised')

- Check year of reference [38] on page 18, lines 42-43

- I acknowledge in this review that it is a study protocol, however feel the authors could have provided some strengths and limitations of the study, particularly the sample size which is quite sizable and how they could use this to compare against Western-based studies, the limitations such as confounders and how contamination of the intervention arm will be minimised should have been stated

- The process evaluation is stated to align with the UK Medical Research Council but no check-list for the main trial was provided |

REVIEWER	Taren Sanders Institute for Positive Psychology and Education, Australian Catholic University
REVIEW RETURNED	26-Jul-2017

| GENERAL COMMENTS | Overall, I commend the authors for their use of the MRC guidelines |

to develop their intervention. It seems they have done significant developmental work to guide the intervention.

I am concerned that the protocol is being submitted for publication after the intervention has been delivered and the first follow-up measures (the primary outcome time point) have been taken. As the intervention has already been delivered, it is not possible to know how the intervention delivery deviated from that which was planned (which is one of the purposes BMJ Open states for publishing protocols).

Sample size calculation – it does not appear that ICCs for class were taken into account? The average class size for Guangzhou is stated at 45, but it does not seem to factor into the power analysis.

Randomisation and blinding – please clarify the stratification procedure. It is stated that there is stratification based on two factors (snacks, indoor activity room), but please clarify that these are balanced across the intervention/control schools? Also, are the data collectors blinded to study allocation? This seems important as some of the data is collected one-on-one with the children.

Measures – Please clarify who is completing which measures, as it is unclear as currently stated. This is especially true of the Health Behaviour Measures, where it is not clear which measures the child completes, and which ones the parents complete.

REVIEWER	Dr. Dorothea Kesztyüs MPH Institute of General Medicine Ulm University Ulm, Germany
REVIEW RETURNED	31-Jul-2017

| GENERAL COMMENTS | Reviewers report
Manuscript ID: bmjopen-2017-018415
Title: A cluster-randomised controlled trial to assess the effectiveness and cost-effectiveness of an obesity prevention programme for Chinese primary school-aged children: the CHIRPY DRAGON study protocol

In this manuscript the authors report a study protocol for an obesity prevention programme in Chinese primary school children. They describe the developmental process including pre-studies and the implementation of the programme. The design of the trial and the respective methods are presented.
Comments and recommendations on the content of the manuscript
o Overall: This is a very well written and elaborated manuscript on an important topic. Although many studies are conducted and reported in this field, the target population and the implementation conditions are of particular interest.
The economic evaluation is extremely important. Yet some aspects should be addressed to improve the manuscript.
Please work thoroughly through your manuscript and correct some typing errors (e.g. page 9, line 34 "send" instead of "sent", line 51 " CHIRPY" instead of "CHIPRY") and some grammatical errors or omissions (e.g. Abstract, line 7: than in other countries; line 11: one of the first examples; line 23: 1,640 pupils; page 14 , line 28: account |

for missing data). I understand that you have to present your abstract in 300 words, but I think you can do this without limiting the readability. For instance in line 14 you could write "methods of the MRC" instead of "methods set out by the MRC" and save two words. Sorry if this seems to be pedantically, but especially for non-native English-speakers readability is essential.

o Background (and general): In the Cochrane review of Langford et al. (1) on the WHO Health Promoting School framework you will find some information on cost-effectiveness and even more information on this topic can be found in the articles of John et al. 2010 (2) and 2012 (3). You state cost-effectiveness as a primary outcome (page 4, line 5 ff.) so you should give some more information.

- Page 3, line 34: A cross-sectional study can only find associations, not causality and grandparents are not yet a validated risk factor. Please differentiate your statement here in terms of known risk factors and factors you identified to be associated with the outcome.

o Methods:
- Method of randomisation and blinding, page 6: Is the social environment of the schools comparable? This could also influence the outcome.

- Comparator, page 11, line 5: Are the on-going health related activities in the control schools assessed? They might compromise the results.

- Outcome measures, page 11, line 10: Just for my own understanding, are there no Chinese growth charts for children available? Why do you use the WHO growth charts?

- Anthropometric measures, page 11, line 45: You should indicate how waist circumference is measured because there are several possible measurement sites (4)

- Health behaviour measures, page 12, line 3: Please describe more in detail how you will use the accelerometer because this is an important topic. How many children do you want to examine, how long will they have to wear them, weekdays and weekends, what kind of data will you measure and how will you analyse the data etc. There is a reference in your reference list (38), but not in the text.

- Other measures, page 12, line 16: The reference number for PedsQL is wrong and I fear most of the other reference numbers, too. You should check all references.

- Other measures, page 12, line 18: Do you use a self-report or a parent-proxy report version of the PedsQL? Children are aged 6-7 years at baseline but PedsQL is for children aged 8-12, what is your rationale? There are two references for the EQ-5D, none is correct. If you included the EQ-5D to calculate QALYs of the children, I don't think the adult version of the EQ-5D is applicable here. The proxy version of the EQ-5D-Y should have been used (https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/). Please state clearly which items are included in the parental questionnaires and which items are administered to the children directly. You should update this paragraph to make it

clearer.

- Economic Evaluation, page 14: This paragraph lacks a lot of information. How will the costs be assessed, retrospectively or alongside the trial? You should mention the perspective here. Is any discounting done and if not, please explain why. Is there a threshold for QALYs in the Chinese population? Or how do you want to decide on the cost-effectiveness of your intervention? The CHU9D is not mentioned in the description of the quality of life measures on page 12. Together with the wrong reference numbers this is somehow confusing and I recommend carefully reworking this part. Incremental cost-effectiveness ratio should be included and explained here. If you want to follow the concept of a piggyback study you will find some information here: https://www.dedipac.eu/policies-and-interventions-toolbox/Piggyback_guideline.pdf.

- You could give some more detailed explanation why you included four different measures of HRQoL (PedsQL, Kidscreen, EQ-5D, and CHU9D).

References
1. Langford R, Bonell CP, Jones HE, Pouliou T, Murphy SM, Waters E, et al. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. Cochrane Database Syst Rev. 2014;4(4):CD008958.

2. John J, Wenig CM, Wolfenstetter SB. Recent economic findings on childhood obesity: cost-of-illness and cost-effectiveness of interventions. Curr Opin Clin Nutr Metab Care. 2010;13(3):305–13.

3. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. Nutrition. 2012;28:829–39.

4. Wang J, Thornton JC, Bari S, Williamson B, Gallagher D, Heymsfield SB, et al. Comparisons of waist circumferences measured at 4 sites. Am J Clin Nutr. 2003;77(2):379–84.

VERSION 1 – AUTHOR RESPONSE

Responses to reviewer 1

1. The anticipated statistical analysis was very sparsely described in the abstract. An additional sentence to elaborate would have been preferred.

Response: We have added in more information on primary and secondary analyses in the abstract.

2. 10/22 studies which showed no, or low drop-out rates which was used to help justify their sample size calculation. The authors should have reviewed the most relevant papers from those 22 studies to identify the optimum drop-out rate or use all 22 studies to inform the expected drop out rate for this study. The authors lacked justification for using a drop-out rate of 10% based on the 10/22 studies which showed low, or no drop-outs.

Response: We have now clarified this in the text on page 14. In fact this review highlighted poor reporting and many common methodological flaws among the included trials. Among 22 included trials, 12 did not report an exact drop-out rate (either not mentioned or reported as "low") and the remainder reported no drop-outs (10/22).

Thus, it was not possible to base the likely drop out estimate on previous studies. We did not assume a 0% drop out, but instead used a conservative estimate of 10% based on unpublished anecdotal evidence in this study setting.

3. Whilst the authors state they sought advice from a medical statistician and local education authorities to identify important factors in which to stratify the schools allocation to intervention or control groups it was unclear why other important factors were not used (i.e. that are associated with obesity e.g. region of Guangzhou [is diet the same or different across this city? Are all schools located within walking distance to the homes of the children attending them? What existing mechanisms are already in place in schools - do some already adhere to some physically active programme? Or have more sports-related activities? Are their extra-curricular activities such as sports available at all schools? Are some of the schools located in more remote, or rural regions of the city? Is socioeconomic status different across this city (e.g. are parents more likely to drive their children in some areas, whereas in other areas they tend to walk), are there other considerations such as the local environment in proximity of the school (are they close to parks or places were physical or sporting events can take place or are they close to fast food outlets, stores which sell unhealthy foods and availability of gaming centers etc). A few are listed as potentially being recorded as potential moderators however, there is a lack of information and detail here. The authors and researchers here could have possibly made the design slightly more complex by applying this as a multi-level cluster randomised control trial.

Response: We thank the reviewer for this detailed comment. All potential major moderators of the intervention effect were considered, including various behavioural, socio-economic and policy conditions of the local families and schools. We sought to include only the most important factors at this design stage to ensure baseline balance. Other factors that contribute to obesity will be adjusted for at the analysis stage. Our research team and partners consist of local residents who work in the city's school health and education authorities. Together with the advice from our medical statistician their in-depth knowledge of the social and policy environment of the study setting informed our randomisation design.

We now have added in extra information on page 6, under the 'Method of randomisation and blinding' section.

In response to the reviewer's specific examples of potential moderators, most of these were not considered to vary sufficiently to be included as factors for stratification. Diet across this city (and within this province and nearby provinces such as Guangxi) is similar (known as the Yue style, one of the 8 distinct styles of Chinese cuisine). In terms of socio-economic and neighbourhood environmental factors, only public (state-funded) schools from the traditional urban districts of this city are eligible for this study. Such eligibility criteria will exclude children of migrant parents who would

require a very different set of interventions. Moreover, in this city, children must follow a standard catchment policy for public primary school registration. This ensured a similar proximity between the home and school. Finally, school regulation and management is highly centralised and standardised in China, especially in relation to school food environment (e.g. no food stores on campus policy) and physical activity provision on campus (e.g. nation-wide standard requirement). Thus, we did not anticipate these factors would differ sufficiently between schools to warrant including them in the stratification strategy.

4. Were birth anthropomorphic measures collected (esp. birth weight which is predictive of later life health and well-being)

Response: Yes, child birth weight will be collected through the parent questionnaire. We have added this to page 11, under 'outcome measures'.

5. A more detailed list of potential confounders that would be adjusted for in the statistical analysis would have been ideal

Response: We added further detail on the covariates that will be included in the secondary analyses on Page 15, under 'planned statistical analysis':

Secondary analyses will additionally adjust for pre-specified school- and child-level covariates. These include those that were used in randomisation (i.e. whether the school provides mid-morning snack, whether the school has an indoor activity room) and important socio-demographic (i.e. sex and parental education level) and health behaviour factors (fruit and vegetables, unhealthy snacks and sugar added drink consumption and minutes/day MVPA and sedentary time).

6. There were no explicit inclusion and exclusion criteria. How could potential sources of contamination be minimised? E.g. siblings, relatives, friends who attend schools in a different treatment group. Is it possible the intervention effects could be diluted?

Response: Regarding our inclusion/exclusion criteria, we have included additional information under 'Study population and participants eligibility' (page 5) and 'Participant recruitment' (page 6). As far as contamination is concerned, we have added in a new section on 'Strategies to minimise contamination' (page 9). Given the presence of the 'One Child Family Planning Policy' in China, siblings are still very rare in this study setting. Moreover, our formative research among the study population (to be published) found that most families are now living in private accommodation. The shift from organisational/employer based residential communities to private housing arrangements has led to poor neighbourhood connectedness, thus making the possibility of contamination low. The new section 'Strategies to minimise contamination' explains the major risks of contamination that we identified and the strategies we will use to minimise those risks.

7. Whilst the school was randomised per cluster, could the detail recorded per student be used for longitudinal analyses (i.e. linked with follow-up data) or will it only be summarised at the school or treatment level?

Response: Individual data will be linked longitudinally, and analysis will be at the individual level, accounting for clustering

8. Spelling error (page 7, line 15, should be CHIRPY DRAGON not CHIRPYD RAGON) and (page 7, line 29 'summaries' should be 'summarised')

Response: We thank the reviewer for this and have corrected these spelling errors on page 7.

9. Check year of reference [38] on page 18, lines 42-43

Response: Thanks for this, we have added the year for this reference.

10. I acknowledge in this review that it is a study protocol, however feel the authors could have provided some strengths and limitations of the study, particularly the sample size which is quite sizable and how they could use this to compare against Western-based studies, the limitations such as confounders and how contamination of the intervention arm will be minimised should have been stated

Response: We have now expanded this section as suggested (pages 16-17).

 11. The process evaluation is stated to align with the UK Medical Research Council but no check-list for the main trial was provided

We have now completed and attached the 2013 SPIRIT checklist. All page numbers refer to those shown on the revised manuscript with track change.

Responses to reviewer 2

1. I am concerned that the protocol is being submitted for publication after the intervention has been delivered and the first follow-up measures (the primary outcome time point) have been taken. As the intervention has already been delivered, it is not possible to know how the intervention delivery deviated from that which was planned (which is one of the purposes BMJ Open states for publishing protocols).

Response: We checked with the editorial team for their policy regarding the timing of RCT protocol submission before the first author went for her maternity leave. We were advised that RCT protocol submissions will be accepted before data collection is completed. Therefore, the manuscript was submitted to the journal before the first follow-up measures were completed. It is possible to know whether/how the intervention delivery deviated from what was planned, because the design of the trial and the intervention programme was fully described within the prospective trial registration, which was published online two years ago.

2. Sample size calculation – it does not appear that ICCs for class were taken into account? The average class size for Guangzhou is stated at 45, but it does not seem to factor into the power analysis.

Response: In this study a cluster is a class, because we will aim to recruit one class (both children and their family members) from each participating school to take part in outcome measures. For simplicity, we have replaced the term 'cluster' by 'school' throughout. We also added the following details to the sample size section (page 14):

'Clustering will be allowed for at the level of the school (in this study one cluster refers to one class because in each participating school, we will aim to randomly select one class to take part in outcome measures). In cases where multiple classes will be included in each school, we will include both levels of clustering where possible'.

3. Randomisation and blinding – please clarify the stratification procedure. It is stated that there is stratification based on two factors (snacks, indoor activity room), but please clarify that these are balanced across the intervention/control schools? Also, are the data collectors blinded to study allocation? This seems important as some of the data is collected one-on-one with the children.

Response: We have added in the following information on page 6. The two factors used for stratification are balanced across the intervention and control schools. We will recruit external workers (independent data collectors who are blinded to study allocation) to undertake all outcome measures.

4. Measures – Please clarify who is completing which measures, as it is unclear as currently stated. This is especially true of the Health Behaviour Measures, where it is not clear which measures the child completes, and which ones the parents complete.

Response: We have added in a table (as a supplement) which summarises the key measurements undertaken in this study. It also specifies what measures are included in the parental, other family member or child questionnaires respectively. The following line was added to the end of the 'data collection methods' paragraph:

'A summary of key measurements undertaken in this study is included in Supplement 1'.

Responses to reviewer 3

1. Please work thoroughly through your manuscript and correct some typing errors (e.g. page 9, line 34 "send" instead of "sent", line 51 " CHIRPY" instead of "CHIPRY") and some grammatical errors or omissions (e.g. Abstract, line 7: than in other countries; line 11: one of the first examples; line 23: 1,640 pupils; page 14 , line 28: account for missing data). I understand that you have to present your abstract in 300 words, but I think you can do this without limiting the readability. For instance in line 14 you could write "methods of the MRC" instead of "methods set out by the MRC" and save two words. Sorry if this seems to be pedantically, but especially for non-native English-speakers readability is essential.

Response: We have made necessary changes as suggested.

2. Background (and general): In the Cochrane review of Langford et al. (1) on the WHO Health Promoting School framework you will find some information on cost-effectiveness and even more information on this topic can be found in the articles of John et al. 2010 (2) and 2012 (3). You state cost-effectiveness as a primary outcome (page 4, line 5 ff.) so you should give some more information.

Response: Thank you for highlighting these important studies. The reviewer is correct to point out that John et al. 2010 and 2012 provide a useful summary of cost-effectiveness of interventions to either prevent or treat childhood obesity but the overwhelming finding from these papers is that it is difficult to draw any conclusions regarding the most cost-effective alternatives due to methodological differences in evaluation approaches and "a better understanding and a more precise assessment of the health care costs and the broader economic burden are necessary". This study will help with both of these claims in that it will provide a very precise and transparent estimation of intervention costs and broader household costs linked to dietary behaviour, and if appropriate, will extrapolate any findings over the course of a lifetime to predict long term costs and benefits associated with preventing childhood obesity – more detail is provided in our response to point 11 below.

3. Page 3, line 34: A cross-sectional study can only find associations, not causality and grandparents are not yet a validated risk factor. Please differentiate your statement here in terms of known risk factors and factors you identified to be associated with the outcome.

Response: This sentence on page 3 has been rewritten as suggested.

4. Method of randomisation and blinding, page 6: Is the social environment of the schools comparable? This could also influence the outcome.

Response: Please see our response to the 3rd comment from reviewer 1.

5. Comparator, page 11, line 5: Are the on-going health related activities in the control schools assessed? They might compromise the results.

Response: Our research partner in China (School Health Unit within the Guangzhou Centre for Disease Control and Prevention) monitors all health-related activities taking place in local primary schools (including schools from both arms of this study). Therefore, throughout the entire trial period, any non-trial related major changes to standard practice (e.g. voluntary introduction of new health promotion programmes and other health intervention trials) in all participating schools will be reported to the research team. We have also explained in response to the 3rd comment from reviewer 1 that school management and regulation in China are highly centralised and standardised, especially in relation to on-campus health promotion activities. Therefore, we do not anticipate this to be an issue.

6. Outcome measures, page 11, line 10: Just for my own understanding, are there no Chinese growth charts for children available? Why do you use the WHO growth charts?

 Response: In a previous cross-sectional study, we examined the agreement (using Kappa), among classifications of children into healthy-weight, overweight and obese categories, by WHO 2007 standards, WGCO reference norm and IOTF. We found very good agreement between the WGCO and WHO classifications (coefficient=0.83, p<0.001). In addition, using the WHO growth charts will facilitate international comparison of trial results. For those reasons, the WHO 2007 standards were chosen for this study.

7. Anthropometric measures, page 11, line 45: You should indicate how waist circumference is measured because there are several possible measurement sites (4)

Response: We have added in extra information on page 11 (i.e. measured midway between the rib cage and the iliac crest) as suggested.

8. Health behaviour measures, page 12, line 3: Please describe more in detail how you will use the accelerometer because this is an important topic. How many children do you want to examine, how long will they have to wear them, weekdays and weekends, what kind of data will you measure and how will you analyse the data etc.

All Children taking part in outcome measures will be invited to wear the wrist-worn accelerometer continuously for 5 days (including 3 weekdays and a weekend). MVPA (minutes/24hours) and sedentary time (minutes/24 hours) will be derived and analysed using the methods described by Noonan and colleagues (Noonan RJ, Boddy LM, Kim Y, Knowles ZR, Fairclough SJ. Comparison of

children's free-living physical activity derived from wrist and hip raw accelerations during the segmented week. J Sports Sci. 2016:1–6. doi: https://doi.org/10.1080/02640414.2016.1255347).

Response: We have added in the above information under Health behaviour measures, on page 12.

9. There is a reference in your reference list (38), but not in the text. Other measures, page 12, line 16: The reference number for PedsQL is wrong and I fear most of the other reference numbers, too. You should check all references.

Response: Thank you very much for spotting these referencing errors. The citation numbers were correct within the text but the order and numbers within the reference list were incorrect. We have corrected these errors.

10. Other measures, page 12, line 18: Do you use a self-report or a parent-proxy report version of the PedsQL? Children are aged 6-7 years at baseline but PedsQL is for children aged 8-12, what is your rationale? There are two references for the EQ-5D, none is correct. If you included the EQ-5D to calculate QALYs of the children, I don't think the adult version of the EQ-5D is applicable here. The proxy version of the EQ-5D-Y should have been used (https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/). Please state clearly which items are included in the parental questionnaires and which items are administered to the children directly. You should update this paragraph to make it clearer.

Response: The reviewer is correct that the PedsQL is recommended for use in children aged 8 years or more. However, we had tested the use of the instrument in 150 children aged 6-7 years as part of a feasibility study, using trained researchers to administer the questionnaire. We found no problems in children's understanding of the questions or difficulties in responding. Furthermore, PedsQl has been used in younger children in other studies as well (https://doi.org/10.1186/s12889-015-1800-8). Although at baseline the children will be aged 6-7 years, they will be close to, at or over 8 years old at both follow-ups. Moreover, this Chinese version of the PedsQL for 8-12 years had been validated (cited in the paper) in our specific study population (Guangzhou city) when we started this study.

The incorrect numbers of the two references have been updated in the reference list.

The Chinese version of EQ-5D we used in this study is for calculating QALYs of the adult family members of participating children. We have clarified this in the paragraph.

We have added in a table (as a supplement) which summarises all the key measurements undertaken in this study. It also specifies what measurements are included in the parental, other family member or child questionnaires respectively.

11. Economic Evaluation, page 14: This paragraph lacks a lot of information. How will the costs be assessed, retrospectively or alongside the trial? You should mention the perspective here. Is any discounting done and if not, please explain why. Is there a threshold for QALYs in the Chinese population? Or how do you want to decide on the cost-effectiveness of your intervention? The CHU9D is not mentioned in the description of the quality of life measures on page 12. Together with the wrong reference numbers this is somehow confusing and I recommend carefully reworking this part. Incremental cost-effectiveness ratio should be included and explained here. If you want to follow the concept of a piggyback study you will find some information here: https://www.dedipac.eu/policies-and-interventions-toolbox/Piggyback_guideline.pdf.

Response: Due to word count restrictions, information on the economic evaluation was deliberately kept brief however on reflection, we agree with the reviewer that more information is required. We therefore have re-written this section in pages 15-16 as follows:

'The economic evaluation will be conducted alongside the trial to estimate the difference in costs and outcomes between the situation of supplying the intervention package and a situation where no intervention is in existence (i.e. usual current practice). Both a cost-utility analysis and a cost-effectiveness analysis (CEA) will be conducted. The primary measure of effectiveness, for the CEA will be change in BMI-z score between the arms. The primary outcome measure for the CUA will be Quality-Adjusted Life Year (QALY) calculated from data collected using the CHU9D instrument, a paediatric utility-based Quality of Life measure validated in a Chinese population. Costs collected will focus on costs that are likely to differ between the intervention and control arm: set up and delivery of intervention; staff costs; materials used during sessions; and impact of intervention upon household food expenditure. The trial facilitators will supply all data on intervention-related costs. Costs linked to household expenditure will be collected from participants asked to complete a parent report. Unit costs will be obtained from Chinese sources, or valued at market prices. The costs will be estimated by calculating the resource use multiplied by the unit cost. Total costs for the intervention as well as average cost per participant will be estimated. Using regression analysis, we will control for differences in characteristics of participants, and baseline CHU9D scores

The economic evaluation will be conducted from both a public sector and societal perspective. The public sector perspective will only include costs linked to delivery of the intervention alongside the average QALY impact on the children; the societal perspective will broaden the framework to include household expenditure and intervention effect on parents and other household adult members. To facilitate this broader perspective, QALYs for parents and other adult household members will be estimated using data collected from the EQ5D instrument. Costs and outcomes will be combined using incremental cost effectiveness ratio (ICER) and expressed as cost per QALY gained. This information can then be used by decision makers to judge the cost-effectiveness. As no equivalent threshold value exists within the Chinese setting for how much decision makers are willing to pay for a unit gain in QALY, the ICER will be assessed with reference to the established- UK and -US threshold values.

Uncertainty in the data will be assessed using non-parametric bootstrapping and decision uncertainty will be represented using cost-effectiveness acceptability curves.

The within-trial analysis will only use data from the trial and if the intervention demonstrates effectiveness within that time period we will develop a decision-analytic model to estimate the cost-effectiveness beyond the trial period. This will project costs and outcomes over a lifetime and probability sensitivity analysis will be applied to reflect the uncertainty within the model, and the data used to populate the model'.

12. You could give some more detailed explanation why you included four different measures of HRQoL (PedsQL, Kidscreen, EQ-5D, and CHU9D).

Response: The economic evaluation will use information collected from the EQ5D and the CHU9D as explained in the response to point 11 above.

PedsQL will be included as it is a widely used measure of quality of life in children, and to allow comparability with other data. However, it does not allow us to derive a measure of utility. Therefore, the CHU9D will also be assessed. Only one domain from the Kidscreen questionnaire (related to social acceptance) will be included. This is because social acceptance is an important outcome when considering children's weight status and is not covered in the other quality of life measurements. As

described previously, the EQ5D measure will be used for assessing quality of life in parents and guardians, and not in the children. This will allow us to derive utility values in parents to assess any wider potential impacts of the intervention

Additional changes made by the authors

In addition to the comments above, whilst revising the paper we noted that the methods section would be improved by moving the sample size section from page 5 to page 14. We have therefore also made this additional amendment.

We also added James Martin to the author list, and indicated his role in the contributions section. James performed randomisation of the trial as an independent statistician and contributed to the revision of the randomisation section.

VERSION 2 – REVIEW

REVIEWER	Dr. Dorothea Kesztyüs MPH
	Instiute of General Medicine, Ulm University, Ulm, Germany
REVIEW RETURNED	25-Sep-2017

| **GENERAL COMMENTS** | The authors have answered all questions and implemented all necessary changes to improve their manuscript. |