

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	EdAI-2 (Educació en Alimentació) programme: reproducibility of a cluster randomised, interventional, primary-school-based study to induce healthier lifestyle activities in children
AUTHORS	Llauradó, Elisabet; Tarro, Lucia; Moríña, David; Queral, Rosa; Giralt, Montse; Solà, Rosa

VERSION 1 - REVIEW

REVIEWER	Gabriel Ángel Martos Moreno Hospital Niño Jesús. Universidad Autónoma de Madrid.
REVIEW RETURNED	11-Jun-2014

GENERAL COMMENTS	<p>Llauradó et al. afford an evaluation of the reproducibility of a previously developed educational school-based intervention program to improve lifestyle, including diet and physical activity over a period of 22 months in 690 pupils 7-8 years at enrollment, (320 in the intervention group and 370 in the control group); focused on 8 lifestyle topics and implemented by university students acting throughout 3 school academic years. Obesity prevalence was considered as the primary outcome measure (no changes observed after intervention in any group) and body mass index, dietary and lifestyle habits (questionnaires) were considered as secondary outcome measures. The authors conclude that their school-based intervention program is feasible, reproducible and adaptable to any school environment, being able to lead to an increased percentage of children (mainly boys) performing ≥ 4 after-school physical activity hours / week.</p> <p>The manuscript needs English revision. The studied topic (school-based intervention programs in obesity prevention) is of interest, due to the current epidemiological impact of childhood obesity.</p> <p>However, the performed study does not achieve its main objectives and focuses on a partial interpretation of the positive post-intervention results yielded, with minimum attention displayed to</p>
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

multiple “negative” results raised in the comparison after the intervention. Also, some methodological considerations must be performed. To address this issue the following comments are raised.

Major comments:

1. When analyzing a population with high variability (sex, ethnic origin and age) standardized values for any auxological measurement rather than raw data should be used (i.e. Table 1). Moreover, the population selected is followed from baseline (7-8 years chronological age) up to 3 academic years (10-11 years of chronological age). According to gender, a significant proportion of the females included will probably have started puberty throughout the follow-up period, thus introducing an additional (non considered) confounding factor in the comparisons established.
2. The population selected to test the reproducibility of the previously designed trial is some 80 km away from the originally studied population and, most probably, made upon a similar ethnic and socio-cultural background. These conditions limit, to a major extent, the conclusion raised by the authors about one of the main objectives of the study: **“Our school-based intervention program is feasible, reproducible and adaptable to any school environment”**. Also, the development of the intervention in every school from a single town (Amposta), with all of the control groups belonging to centers from surrounding cities, as well as the condition imposed of a 50% minimum involvement of the classroom pupils to implement the intervention, strongly impair aleatorization thus establishing populational biases prior to the comparison.
3. Regarding the main outcome measurement proposed. No differences in obesity prevalence in intra- nor in inter-groups comparisons were observed. In contrast, it was striking that, whereas the Control group did not show differences in overweight prevalence either, the intervention group showed a significant increase in overweight prevalence in the whole cohort and, more intensely, in females (Table 2). These data are opposite to the proposed hypothesis of a health promoting effect

	<p>of the proposed program. Interestingly, this trend was reproduced when BMI-SDS evolution throughout the follow-up was investigated (increased in the whole intervention cohort and in females in both groups, Table 3). This increase in BMI-SDS in both groups could be, at least in part, biased by the eventual role of puberty onset and progression (increase in fat mass) in females along the follow-up.</p> <p>4. The authors remark the observed significant difference in the increase, <u>exclusively</u> (not mainly as concluded in the abstract) in boys performing >4 hours physical activity in the intervention group regarding the control one (Table 4). However, in the same table, higher “improvements” or lower “worsening” in the percentages of TV watching in boys are observed in the control group regarding the intervention one (ranges <2 and 2 to 4 hours/day) are reported. Similarly, it is shown a higher decrease in pastry consumption in control boys or the higher increase in fast-food consumption in girls in the intervention group when compared to the control one. All of these results should be considered as a whole, as there are a higher number of “pros” in the control group than in the intervention one.</p> <p><u>Minor comments:</u></p> <ol style="list-style-type: none"> 1. Name, gender, date and place of birth and informed consent sign up cannot be considered as inclusion criteria (Abstract & Methods sections). 2. The proportion of Latin American patients could be significantly different between groups. The authors should check whether there were any differences in ethnic distribution between groups. 3. Chronological age should be included in Table 1 for every group at both time-points. 4. Reference format should be checked and standardized..
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Fernando Vio
----------	--------------

	Institute of Nutrition and Food Technology (INTA) University of Chile
	CHILE
REVIEW RETURNED	26-Jun-2014

GENERAL COMMENTS	<p>1. In Results, there is not a Table with Food Habits and Food Consumption. The results are well described, but a Table N° 4 would be necessary.</p> <p>2. Please explain better the reduction in TV watching, because in Table 4. there is just a non significant reduction in >4h/day.</p> <p>3. There are not References related to Latin American Publications related to the same topic, such as the list below, at the end of the report.</p> <p>4. Discussion should be more exhaustive related to international literature, not only focused in KOPS, Pro Children and ALADINO studies.</p> <p>An interesting and well written article in a fundamental topic in public health in the world, which is the prevention of children obesity with nutrition education and physical activity intervention at basic schools. It requires some modifications in Results:</p> <ol style="list-style-type: none"> 1. To include a Table with Food Habits and Food Consumption 2. To explain better the reduction of TV watching (for me was not so clear the reduction) 3. To include in the Discussion and References other experiences, besides the KOPS, Pro Children and ALADINO studies. <p>1. Lobos L, Leyton B, Kain J, Vio F. Evaluación de una intervención educativa para la prevención de la obesidad infantil en escuelas públicas de Chile. <i>Nutr Hosp</i> 2013; 28:1156-1164.</p> <p>2. Kain J, Leyton B, Concha F, Weisstaub SG, Lobos L, Bustos N, Vio F. Evaluación de una intervención en educación alimentaria y actividad física para prevenir obesidad infantil en escuelas públicas de Santiago de Chile. <i>Arch Latinoam Nutr</i> 2012; 62:60-67.</p> <p>3. Vio F, Zacarías I, Lera L, Benavides MC, Gutiérrez AM. Prevención de la obesidad en escuelas básicas de Peñalolén: Componente alimentación y nutrición. <i>Rev Chil Nutr</i> 2011; 38:268-276.</p> <p>4. Ratner R, Durán S, Garrido MJ, Balmaceda S, Jadue L, Atalah E. Impacto de una intervención en alimentación y actividad física sobre la prevalencia de obesidad en escolares. <i>Nutr Hosp</i> 2013; 28:1508-14</p> <p>5. Bacardí-Gascon M, Pérez-Morales ME, Jiménez-Cruz A. A six month randomized school intervention and an 18-month follow-up intervention to prevent childhood obesity in Mexican elementary schools. <i>Nutr Hosp</i> 2012; 27:755-62.</p> <p>6. Aguilar MJ, González E, García CJ, García PA, Álvarez J, Padilla CA, González JL, Ocete E. Obesidad de una población de escolares de Granada: evaluación de la eficacia de una intervención educativa. <i>Nutr Hosp</i> 2011; 26:636-64.</p> <p>7. Rauh C, Kovalskys I, De Gregorio MJ. Gender differences and a school-based obesity prevention program in Argentina: a randomized trial. <i>Rev Panam Salud Publica</i> 2013; 34:75-82.</p> <p>8. Ruiz Santana S, Ruiz Hernández F, Hernández CR, Álvarez A, Saavedra P. Evaluación de dos intervenciones nutricionales y de hábitos de higiene y salud en una cohorte de alumnos de enseñanza primaria. <i>Nutr Hosp</i> 2009; 24:596-606.</p>
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>9. Espejo MP, Vásquez MD, Benedi V, López JC. Hábitos de alimentación y de actividad física. Un año de intervención en la escuela. Estudio piloto. Acta Pediatr Esp 2009; 67:21-25.</p> <p>10. Vio F, Salinas J, Montenegro E, González CG, Lera L. Efecto de una intervención educativa en alimentación saludable en profesores y niños preescolares y escolares de la Región de Valparaíso, Chile. Nutr Hosp 2014; 29: 1298-1304.</p> <p>11. Kain J, Uauy R, Concha F, Leyton B, Bustos N, Salazar G, Lobos L, Vio F. School-based obesity prevention interventions for Chilean children during the past decades: Lessons learned. Adv Nutr 2012; 3(4):616s-621s.</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Roya Kelishadi,MD Professor of Pediatrics, Faculty of Medicine Isfahan University of Medical Sciences, Isfahan, Iran
REVIEW RETURNED	30-Jun-2014

GENERAL COMMENTS	<p>I think the paper should be rejected.</p> <p>The study does not add anything to the current literature. It provides well-known facts about childhood obesity and school-based interventions. Almost all such interventions are effective during the trial, but are no more continued after the trial.</p> <p>The current protocol does not present any specific item to the current interventions.</p> <p>The English writing needs to be much improved.</p>
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Jake Olivier School of Mathematics and Statistics University of New South Wales Sydney, Australia
REVIEW RETURNED	05-Aug-2014

GENERAL COMMENTS	<p>I found the study's goals, i.e., to assess the reproducibility of an RCT, to be interesting and worth publication. However, I found much of the manuscript confusing and not presented adequately. Specific comments follow.</p> <p>Treatment allocation to schools seem to based on geography (intervention schools are in Amposta while control schools are in surrounding towns). How exactly is this a randomised controlled trial? Also see pages 5-6, lines 84-85.</p> <p>Inclusion criteria includes name, gender and DOB. Are those missing that information excluded or were some people excluded based on these criteria? If so, why were those with missing names excluded? It doesn't seem as though it would affect the analysis.</p> <p>Page 3, line 24. Odds ratios cannot be negative (a similar error is found in the results). Please recheck your statistical results.</p> <p>Some abbreviations are not defined when they first appear in the text, for example, OB in the abstract</p> <p>Page 5, line 60. Which Korea is being mentioned here? I assume it's South Korea.</p>
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>Secondary schools are listed in the abstract, yet it appears only primary schools were chosen (page 5, line 83). Please clarify.</p> <p>Page 8, line 166. It's more correct to say the analysis was 'stratified' by gender and not segregated.</p> <p>Page 9, line 169. How are BMI z-scores computed? Are they based on summary statistics for the overall sample, stratified by gender, or using population values?</p> <p>Please provide more information about computing sample size. There are over 60 methods for sample size comparing two proportions.</p> <p>For the linear mixed models, was the outcome proportions of obese? This would assume the proportions are normally distributed. Is that reasonable here?</p> <p>It is unclear to me how McNemar's test is used on behavioural changes. Were these values dichotomous and, if so, how?</p> <p>The authors seem to be assuming data is missing at random. This is not the same as an "implication" or "understood to be".</p> <p>Tables 2-4. It is unclear to me what is being tested here. Specifically, how is McNemar's test and Fisher's exact test being used?</p> <p>The manuscript could use a thorough revision.</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Gabriel Ángel Martos Moreno

Institution and Country: Hospital Niño Jesús, Universidad Autónoma de Madrid

General: Llauradó et al. afford an evaluation of the reproducibility of a previously developed educational school-based intervention program to improve lifestyle, including diet and physical activity over a period of 22 months in 690 pupils 7-8 years at enrollment, (320 in the intervention group and 370 in the control group); focused on 8 lifestyle topics and implemented by university students acting throughout 3 school academic years. Obesity prevalence was considered as the primary outcome measure (no changes observed after intervention in any group) and body mass index, dietary and lifestyle habits (questionnaires) were considered as secondary outcome measures. The authors conclude that their school-based intervention program is feasible, reproducible and adaptable to any school environment, being able to lead to an increased percentage of children (mainly boys) performing ≥ 4 after-school physical activity hours / week.

The manuscript needs English revision. The studied topic (school-based intervention programs in obesity prevention) is of interest, due to the current epidemiological impact of childhood obesity. However, the performed study does not achieve its main objectives and focuses on a partial interpretation of the positive post-intervention results yielded, with minimum attention displayed to multiple "negative" results raised in the comparison after the intervention. Also, some methodological considerations must be performed. To address this issue the following comments are raised.

Major comments:

Point 1: When analyzing a population with high variability (sex, ethnic origin and age) standardized

values for any axiological measurement rather than raw data should be used (i.e. Table 1). Moreover, the population selected is followed from baseline (7-8 years chronological age) up to 3 academic years (10-11 years of chronological age). According to gender, a significant proportion of the females included will probably have started puberty throughout the follow-up period, thus introducing an additional (non considered confounding factor in the comparisons established)

Response 1: Puberty could have been a limitation of the study since we did not seek this information from the parents or guardians of the participants. We have added this limitation in the revised Discussion Pg 28, Ln 415: "Finally, another limitation could be the proportion of females who may have started puberty in the course of the study. This implies changes in body composition. However, both study groups (intervention and control) had a similar proportion of females with a similar age, and this could cancel-out the effect."

Point 2A: The population selected to test the reproducibility of the previously designed trial is some 80 km away from the originally studied population and, most probably, made upon a similar ethnic and socio-cultural background. These conditions limit, to a major extent, the conclusion raised by the authors about one of the main objectives of the study: "Our school-based intervention program is feasible, reproducible and adaptable to any school environment".

Response 2A: As the reviewer correctly highlights, this area of "Terres de l'Ebre" where the study was conducted is different from the city of Reus and its environs where the original EdAI study was performed. The villages and small towns have different cultural traditions and social characteristics that pertain to a semi-rural environment. However, they are not completely different; the many similarities are common to all communities in this area of Catalunya. The majority of European countries have minority populations that account for <20% of the overall population (Wolff S., 2008). In our case, this is around 20-25%, which enables us to make comparisons with other European countries.

Nevertheless, we have changed the main conclusion:

In the Abstract Pg 3, Ln 25: "Conclusions: Our school-based intervention is feasible, adaptable to quite different school environments, and reproducible. The main improvement was after-school PA (≥ 4 h/week) in boys. Further, TV watching decreased to <2 TV h/day. This suggests that our intervention programme induces healthy lifestyle effects (such as more exercise and less sedentary behaviour) which can produce anti-obesity benefits in children."

In the main text Conclusion Pg 29, Ln 425: "Our school-based intervention is feasible, adaptable to quite different school environments, and reproducible. The main improvement was after-school PA (≥ 4 h/week) in boys. Further, TV watching decreased to <2 TV h/day. This suggests that our intervention programme induces healthy lifestyle effects (such as more exercise and less sedentary behaviour) which can produce anti-obesity benefits in children.."

References:

- Wolff S. Ethnic minorities in Europe: The Basic Facts. Centre for International Crisis Management and Conflict Resolution. University of Nottingham, England. 2008.

Point 2B: Also, the development of the intervention in every school from a single town (Amposta), with all of the control groups belonging to centers from surrounding cities, as well as the condition imposed of a 50% minimum involvement of the classroom pupils to implement the intervention, strongly impair aleatorization thus establishing populational biases prior to the comparison.

Response 2B: The reason for using this criterion of the schools themselves as "intervention" or "control" group is because schools situated within the same town have good relationships and communications between them which would encourage "contamination" between the groups in the study. For example, if a school in the control group acquired some information deployed in the intervention group as a result of casual dialogue among parents and school friends, this can motivate control students and families to adopt the intervention recommendations and, as a consequence, do

not behave as controls ought.

In our randomised study, the participants are the towns and, hence, could be termed “cluster” randomisation controlled trial. The towns are allocated at random without taking any similarities or differences between them into account. This is explained in the glossary of the National Institute for Health and Care Excellence (NICE). The defining feature of a cluster RCT is that the unit of allocation is a group rather than an individual (Murray. 2001). The cluster RCT is used for 4 public health reasons (Osrin et al., 2009):

- many public health interventions are delivered to groups, areas, institutions or systems collectively rather than individually;
- testing the delivery of an intervention to an individual might raise concerns if others are not included;
- individually-delivered interventions might spread among family, friends or the wider community;
- they are interested in the mass effectiveness of interventions that will be rolled out to the public with varying degrees of quality, uptake, adherence and response, even if they have been shown to be efficacious in individual cases.

References:

- National Institute for Health and Care Excellence (NICE). Accessed 21st of August 2014 <http://www.nice.org.uk/glossary?letter=R>
- Murray DM. Statistical models appropriate for designs often used in group-randomized trials. *Stat Med* 2001; 20: 1373-85 doi: 10.1002/sim.675 pmid: 11343359.
- Osrin D, Kishwar A, Fernandez A, et al. Ethical challenges in cluster randomized controlled trials: experiences from public health interventions in Africa and Asia. *Bulletin of the World Health Organization* 2009;87:772-779. doi: 10.2471/BLT.08.051060.

We have introduced the terminology in the revised manuscript in Material and methods: study population, Pg 7, Ln 116, “developed a cluster randomisation scheme”

Also, we have added as a limitation in the Discussion, Pg 28, Ln 403 the statement: “Secondly, assigning control groups according to towns surrounding the intervention town could be a limitation. However, schools of the same town have good relationships and communications with each other and this could entail a possible contamination between schools if assigned to intervention or control status within the same town. This cross-contamination would be minimised if the schools themselves were assigned to intervention or control.”

The requirement of a 50% minimum involvement of the overall classroom population for it to be considered for intervention is a characteristic derived from the original EdAI study protocol (Giralt et al, 2011).

Point 3: Regarding the main outcome measurement proposed. No differences in obesity prevalence in intra- nor in inter-groups comparisons were observed. In contrast, it was striking that, whereas the Control group did not show differences in overweight prevalence either, the intervention group showed a significant increase in overweight prevalence in the whole cohort and, more intensely, in females (Table 2). These data are opposite to the proposed hypothesis of a health promoting effect of the proposed program. Interestingly, this trend was reproduced when BMI-SDS evolution throughout the follow-up was investigated (increased in the whole intervention cohort and in females in both groups, Table 3). This increase in BMI-SDS in both groups could be, at least in part, biased by the eventual role of puberty onset and progression (increase in fat mass) in females along the follow up.

Response 3: This is a point of considerable interest. A sensitivity analysis including BMI z-score in the mixed model of repeated measures has been performed for each group and stratified by sex. The results were consistent with those presented in Tables 2 and 3 in which a non-significant effect of z-score BMI was observed throughout.

We have included a sentence in Results Pg 16, Ln 260: “Despite no differences in BMI z-score, the boys of intervention group did not have an increase in percentage fat mass (19.96% to 20.02%; $p=0.896$), whereas intervention girls (22.06% to 23.55%; $p<0.001$), together with boys (19.18% to

20.64%, $p < 0.001$) and girls (23.26% to 24.98%) of control group had a significant increase.”and in Discussion Pg 25, Ln 336: “Despite that OW and OB remained similar between groups, we observed percentage fat mass maintenance in the boys of the intervention group, whereas girls of the intervention and control group had increases.”

Point 4: The authors remark the observed significant difference in the increase, exclusively (not mainly as concluded in the abstract) in boys performing >4 hours physical activity in the intervention group regarding the control one (Table 4). However, in the same table, higher “improvements” or lower “worsening” in the percentages of TV watching in boys are observed in the control group regarding the intervention one (ranges <2 and 2 to 4 hours/day) are reported. Similarly, it is shown a higher decrease in pastry consumption in control boys or the higher increase in fast-food consumption in girls in the intervention group when compared to the control one. All of these results should be considered as a whole, as there are a higher number of “pros” in the control group than in the intervention one.

Response 4: Firstly, in Table 4, now Table 5 we have been alerted regarding the errors and have rectified the p-values. Further, we have re-calculated the TV category, as shown below. In the lowest category of TV h/day (<2 TV h/day), or the change from >2 TV h/day to <2 TV h/day category, there are 43% in the intervention group and 26.5% in the control group that continued to track with assignment group. Hence, a total of 16.5% more of boys of the intervention group, compared to control group, changed to a healthy lifestyle away from TV watching i.e. from a sedentary behaviour ($p=0.009$).

Baseline End-of-study Codifying the “TV Hours” Variable

3-4h/d TV or >4 h/d TV 0-2h/d TV 1 (reduction of TV hours)

0-2h/d TV 0-2h/d TV 1 (no change to the best category of TV hours)

3-4h/d TV or >4 h/d TV 3-4h/d TV or >4 h/d TV 0 (no change to low category of TV hours)

0-2h/d TV 3-4h/d TV or >4 h/d TV 0 (increase in TV hours)

We used the same approach in Physical Activity (PA) categories.

Baseline End-of-study Codifying the “Physical Activity (PA)” Variable

3-4h/w PA or >4 h/w PA 0-2h/w PA 1 (reduction in PA hours)

0-2h/w PA 0-2h/w PA 1 (no change to the best category of PA hours)

3-4h/w PA or >4 h/w PA 3-4h/w PA or >4 h/w PA 0 (no change to low category of PA hours)

0-2h/w PA 3-4h/w PA or >4 h/w PA 0 (increase in PA hours)

We think that these results need to be highlighted because, when analysing habits, it is as important to highlight the improvements as measured as “maintenance” or “increase” in the participants to the “best” category or reduction of participants to the “worse” category such as, for example, reduction in TV hours or increase in after-school PA hours.

We have included an explanatory statement in the Methodology, Outcomes Pg 10, Ln 184: “To obtain a measurement of overall improvement in lifestyle we generated variables such as the maintenance of status in each category as well as the status in relation to changes in each category over the 22 month period.

We included a new sentence in Results, Lifestyles evaluation Pg 22, Ln 306: “At 22 months, the percentage of boys of the intervention group who performed ≥ 4 h after-school PA/week was increased by 15% ($p=0.027$) while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p < 0.009$). The results indicate less sedentary behaviour in intervention than control individuals.”.

Also, we have revised the p-values of Table 4 (now Table 5) with respect to TV and PA differences at

baseline and end-of-study between intervention and control groups.

We have clarified the statement in the Results, and rectified the error regarding reduction in percentage of TV hours in the Lifestyle evaluation section, Pg 22, Ln 305: "Table 5 summarises the time spent in after-school PA, watching TV, playing video games, and other leisure-time activities. At 22 months, the percentage of boys of the intervention group who performed ≥ 4 h after-school PA/week was increased by 15% ($p=0.027$) while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p<0.009$). The results indicate less sedentary behaviour in intervention than control individuals." In the Abstract Pg 3, Ln 19 we have added: "At 22 months, the percentage of boys in the intervention group who performed ≥ 4 h after-school PA h/week was 15% higher ($p=0.027$), while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p=0.009$), compared to controls."

In Table 4, we have provided a broad view of food habit changes and, from which, it can be observed that the most favourable changes appeared in intervention group compared to controls, except in the increase in fast-food consumption.

Minor comments:

Point 5: Name, gender, date and place of birth and informed consent sign up cannot be considered as inclusion criteria (Abstract & Methods sections).

Response 5: We agree. However, we used the same approach as in the original study carried-out in Reus and surrounding areas. We used these criteria to ensure follow-up of these participants beyond the cessation of the original intervention. The study was offered to children between 7-8 years old who attended in the participating schools. If we did not have the name, gender, date and place of birth and the informed consent of their parents or guardians, it precluded the data of the children being included in the study and, hence, they represent "inclusion criteria"

Point 6: The proportion of Latin American patients could be significantly different between groups. The authors should check whether there were any differences in ethnic distribution between groups.

Response 6: We have checked the ethnic distributions, and the only ethnic minority that were represented in the groups are those from Latin American; 10.3% in intervention and 3.5% in control groups ($p<0.001$) We have added this in the Results, Enrolment section Pg 12, Ln 226: "At baseline, there was a significant difference in the distribution with respect to Latin American children (10.3% in intervention and 3.5% in control group; $p<0.001$). The distribution was random. Of note is that there were no significant differences in distributions of OB and/or OW. Also, no differences were observed in terms of response to the intervention in relation to ethnicity".

However, we have added the difference in ethnic distribution as a limitation in Discussion, Pg 28, Ln 408: "Thirdly, the significant difference in Latin American ethnicity between the two groups of the study at baseline could be a limitation. However, there were no significant differences in distributions of OB and/or OW. Also, no differences were observed in terms of response to the intervention study in relation to ethnicity."

Differences in ethnic distributions of participants in the study at baseline

p-value

Western European 0.711

Eastern Europe 0.149

Latin American <0.001

North African Arab 0.113

Point 7: Chronological age should be included in Table 1 for every group at both time-points.

Response 7: OK. We have included the age variable in Table 1 for intervention and control groups at baseline. The age at the end of the study was included in Results, Enrolment Pg11, Ln 218: ". At 22 months, the mean age was 9.67 (95%CI: 9.60, 9.73) in the intervention group (9.68 years in boys and 9.65 years in girls) and 9.86 (95%CI: 9.79, 9.91) in the control group (9.85 years in boys and 9.84

years in girls). The differences in age were not significant in relation to gender.”

Point 8: Reference format should be checked and standardized.

Response 8: We apologise. We have revised the reference section according to the style requirement of the BMJ Open.

Reviewer: 2

Reviewer Name Fernando Vio

Institution and Country Institute of Nutrition and Food Technology (INTA) University of Chile CHILE

General: An interesting and well written article in a fundamental topic in public health in the world, which is the prevention of children obesity with nutrition education and physical activity intervention at basic schools. It requires some modifications in Results.

Response: Thanks for the good opinion overall.

Point 1: In Results, there is not a Table with Food Habits and Food Consumption. The results are well described, but a Table N° 4 would be necessary. (To include a Table with Food Habits and Food Consumption

Response 1: We agree. We have added this information in a new Table 4 and a new Table 5 now contains TV hours and physical activity data. We improved and extended the results describing food habits and consumption in Results, Lifestyles evaluation, Pg 18, Ln 288: “After 22 months of the study, there were 19.7%, 11.2% and 8.2% more girls in the intervention group who consumed a second fruit per day, one vegetable per day and fast-food weekly than the girls of control group ($p < 0.001$, $p = 0.017$ and $p = 0.013$; respectively). However, there were 17.9% and 17.8% more boys in the intervention group who consumed pastry at breakfast and more than one vegetable a day, compared to boys of control group ($p = 0.002$, $p = 0.001$; respectively). Conversely, there were 12.9% and 12.2% more girls in the control group who consumed legumes and cereal breakfast than girls of intervention group ($p = 0.013$, $p = 0.032$; respectively) (Table 4).”

We have widened the Discussion on healthy habits Pg 27, Ln 389: “In the dietary habits aspect of EdAI-2 study, we observed that the increase in healthy lifestyle habits such as the increase in fruit and vegetables consumption and increasing PA h/week while maintaining low TV h/d, are promising lifestyle changes that could induce a reduction of OW and OB over the long-term.”

We have added a statement in Discussion Limitation Pg 28, Ln 412: “Fourthly, when asked about fast-food consumption, the participants interpreted this as pertaining only to fast-food outlets such a burger shops, and did consider other concepts such as frozen pizza consumed at home.”

Point 2: Please explain better the reduction in TV watching, because in Table 4. there is just a non significant reduction in $>4\text{h/day}$. (To explain better the reduction of TV watching (for me was not so clear the reduction)

Response 2: In Table 4 (now Table 5) we have rectified the p-values and, as well, have calculated the change in TV viewing (see below). The lowest category of TV h/day ($<2\text{ TV h/day}$) and/or change from $>2\text{ TV h/day}$ to $<2\text{ TV h/day}$, contained 43% of the intervention group and 26.5% of the control group. Therefore, a total of 16.6% more boys of the intervention group changed to a healthy lifestyle with respect to TV-watching behaviour compared to the control group ($p = 0.009$) i.e. indicating less sedentary lifestyle in the intervention group

Baseline End-of-study Codifying the “TV hours” variable

3-4h/d TV or $>4\text{h/d TV}$ 0-2h/d TV 1 (reduction of TV hours)

0-2h/d TV 0-2h/d TV 1 (no change in the best category hours)
3-4h/d TV or >4h/d TV 3-4h/d TV or >4h/d TV 0 (no change in low category)
0-2h/d TV 3-4h/d TV or >4h/d TV 0 (increase in TV hours)

Also, we applied this approach to the Physical Activity variable.

Baseline End-of-study Codifying the “Physical Activity” variable
3-4h/w PA or >4h/w PA 0-2h/w PA 1 (reduction of PA hours)
0-2h/w PA 0-2h/w PA 1 (no change to the best PA category)
3-4h/w PA or >4h/w PA 3-4h/w PA or >4h/w PA 0 (no change to PA category)
0-2h/w PA 3-4h/w PA or >4h/w PA 0 (increase PA hours)

We think that these results need to be highlighted because, when analysing habits, it is as important to highlight the improvements as measured as “maintenance” or “increase” in the participants to the “best” category or reduction of participants to the “worse” category such as, for example, reduction in TV hours or increase in after-school PA hours.

We have included an explanatory statement in the Methodology, Outcomes Pg 10, Ln 184: “To obtain a measurement of overall improvement in lifestyle we generated variables such as the maintenance of status in each category as well as the status in relation to changes in each category over the 22 month period.

We included a new sentence in Results, Lifestyles evaluation Pg 22, Ln 306: “At 22 months, the percentage of boys of the intervention group who performed ≥ 4 h after-school PA/week was increased by 15% ($p=0.027$) while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p<0.009$). The results indicate less sedentary behaviour in intervention than control individuals.” Also, we have revised the p-values of Table 4 (now Table 5) with respect to TV and PA differences at baseline and end-of-study between intervention and control groups.

We have clarified the statement in the Results, and rectified the error regarding reduction in percentage of TV hours in the Lifestyle evaluation section, Pg 22, Ln 305: “Table 5 summarises the time spent in after-school PA, watching TV, playing video games, and other leisure-time activities. At 22 months, the percentage of boys of the intervention group who performed ≥ 4 h after-school PA/week was increased by 15% ($p=0.027$) while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p<0.009$). The results indicate less sedentary behaviour in intervention than control individuals.” In the Abstract Pg 3, Ln 19 we have added: “At 22 months, the percentage of boys in the intervention group who performed ≥ 4 h after-school PA h/week was 15% higher ($p=0.027$), while there was 16.6% more boys in the intervention group watching ≤ 2 hTV/day ($p=0.009$), compared to controls.”

In Table 4, we have provided a broad view of food habit changes and, from which, it can be observed that the most favourable changes appeared in intervention group compared to controls, except in the increase in fast-food consumption.

Point 3: There are not References related to Latin American Publications related to the same topic, such as the list below, at the end of the report.

Response 3: New references have been added in the revised manuscript in the Background and a comment is added in the Discussion, as stated in Point 4. In Pg 5, Ln 62 “The school is an ideal place for the promotion of healthy nutrition and lifestyle habits,[12] and, as some studies have shown, such interventions have inspired changes in nutritional habits and BMI status,[13-14]; the message being received by all schoolchildren, irrespective of ethnic and socioeconomic differences.[9].” We have added the following references.

13. Vio F, Zacarías I, Lera L, et al. Prevención de la obesidad en escuelas básicas de Peñalolén: Componente alimentación y nutrición. Rev Chil Nutr 2011; 38:268-276.

14. Bacardí-Gascon M, Pérez-Morales ME, Jiménez-Cruz A. A six month randomized school intervention and an 18-month follow-up intervention to prevent childhood obesity in Mexican

elementary schools. *Nutr Hosp* 2012; 27:755-62.

Point 4: Discussion should be more exhaustive related to international literature, not only focused in KOPS, Pro Children and ALADINO studies. (To include in the Discussion and References other experiences, besides the KOPS, Pro Children and ALADINO studies.)

Response 4: The Discussion was focused mainly on studies evaluating reproducibility of an intervention, or on similar school-based intervention to improve healthy lifestyle with a view to reducing or stopping the trend towards obesity. However, in view of this Reviewer's very valid comments, we have incorporated the new references recommended, as well as their relevant messages in Discussion section, Pg 25, Ln 340: "As proposed by Kain J et al, designing a new school-based intervention study needs to have some critical aspects considered. These include: the random allocation of schools, although methodologically desirable, is not always possible; participation of parents is very limited; obesity is not recognised as a problem; and increasing physical activity and implementing training programmes for teachers is difficult due to an inflexible curriculum and lack of teachers' time. Unless these barriers are overcome, obesity prevention programmes will not produce positive and lasting outcomes.[27]. As such, our programme of HPA-implemented intervention activities in classrooms is an attractive alternative that circumvents lack-of-teacher-time. "

References:

27. Kain J, Uauy R, Concha F, et al. School-based obesity prevention interventions for Chilean children during the past decades: Lessons learned. *Adv Nutr* 2012; 3(4):616s-621s.

References proposed by reviewer 2:

1. Lobos L, Leyton B, Kain J, Vio F. Evaluación de una intervención educativa para la prevención de la obesidad infantil en escuelas públicas de Chile. *Nutr Hosp* 2013; 28:1156-1164.
2. Kain J, Leyton B, Concha F, Weisstaub SG, Lobos L, Bustos N, Vio F. Evaluación de una intervención en educación alimentaria y actividad física para prevenir obesidad infantil en escuelas públicas de Santiago de Chile. *Arch Latinoam Nutr* 2012; 62:60-67.
3. Vio F, Zacarías I, Lera L, Benavides MC, Gutiérrez AM. Prevención de la obesidad en escuelas básicas de Peñalolén: Componente alimentación y nutrición. *Rev Chil Nutr* 2011; 38:268-276.
4. Ratner R, Durán S, Garrido MJ, Balmaceda S, Jadue L, Atalah E. Impacto de una intervención en alimentación y actividad física sobre la prevalencia de obesidad en escolares. *Nutr Hosp* 2013; 28:1508-14
5. Bacardí-Gascon M, Pérez-Morales ME, Jiménez-Cruz A. A six month randomized school intervention and an 18-month follow-up intervention to prevent childhood obesity in Mexican elementary schools. *Nutr Hosp* 2012; 27:755-62.
6. Aguilar MJ, González E, García CJ, García PA, Álvarez J, Padilla CA, González JL, Ocete E. Obesidad de una población de escolares de Granada: evaluación de la eficacia de una intervención educativa. *Nutr Hosp* 2011; 26:636-64.
7. Raush C, Kovalskys I, De Gregorio MJ. Gender differences and a school-based obesity prevention program in Argentina: a randomized trial. *Rev Panam Salud Publica* 2013; 34:75-82.
8. Ruiz Santana S, Ruiz Hernández F, Hernández CR, Álvarez A, Saavedra P. Evaluación de dos intervenciones nutricionales y de hábitos de higiene y salud en una cohorte de alumnos de enseñanza primaria. *Nutr Hosp* 2009; 24:596- 606.
9. Espejo MP, Vásquez MD, Benedi V, López JC. Hábitos de alimentación y de actividad física. Un año de intervención en la escuela. Estudio piloto. *Acta Pediatr Esp* 2009; 67:21-25.
10. Vio F, Salinas J, Montenegro E, González CG, Lera L. Efecto de una intervención educativa en alimentación saludable en profesores y niños preescolares y escolares de la Región de Valparaíso, Chile. *Nutr Hosp* 2014; 29: 1298-1304.
11. Kain J, Uauy R, Concha F, Leyton B, Bustos N, Salazar G, Lobos L, Vio F. School-based obesity prevention interventions for Chilean children during the past decades: Lessons learned. *Adv Nutr* 2012; 3(4):616s-621s.

Reviewer: 3

Reviewer Name Roya Kelishadi,MD

Institution and Country: Professor of Pediatrics, Faculty of Medicine. Isfahan University of Medical Sciences, Isfahan,Iran

General: I think the paper should be rejected.

The study does not add anything to the current literature. It provides well-known facts about childhood obesity and school-based interventions. Almost all such interventions are effective during the trial, but are no more continued after the trial.

Point 1: The current protocol does not present any specific item to the current interventions.

Response 1: We beg to differ. The results have added information about healthy lifestyle habits. As described in Discussion, Pg 26, Ln 349: "The EdAI-2 program confirmed that after-school PA (in terms of h/week) can be stimulated in primary school as part of a healthy lifestyle. As we had observed in the original EdAL program,[18] at 28 months of intervention, there was an increase of up to 19.7% of children dedicating >5 hours/week to extra-curricular physical activities,[18]. Further, the after-school PA was maintained despite cessation of the intervention program,[28]. The effect of EdAI program during its implementation and after the official cessation indicated an impact on PA, whereas modification towards healthy food choices occurred according to the site of the program's implementation, and was not consistent."

Point 2: The English writing needs to be much improved.

Response 2: The manuscript has been systematically revised and, we hope, all anomalies of language have been removed.

Reviewer: 4

Reviewer Name: Name Jake Olivier

Institution and Country: School of Mathematics and Statistics. University of New South Wales. Sydney, Australia

General: I found the study's goals, i.e., to assess the reproducibility of an RCT, to be interesting and worth publication. However, I found much of the manuscript confusing and not presented adequately. Specific comments follow. The manuscript could use a thorough revision.

Response: Thanks for the good opinion overall.

Point 1: Treatment allocation to schools seems to be based on geography (intervention schools are in Amposta while control schools are in surrounding towns). How exactly is this a randomised controlled trial? Also see pages 5-6, lines 84-85.

Response 1: A randomised controlled trial (RCT) in which the town (in the current case) forms the unit of randomisation can be termed a cluster randomisation controlled trial, as described in the glossary of the National Institute for Health and Care Excellence (NICE). The defining feature of a cluster RCT is that the unit of allocation is a group rather than an individual (Murray 2001). The cluster RCT is used for 4 public health reasons (Osirin et al., 2009):

- many public health interventions are delivered to groups, areas, institutions or systems collectively rather than individually;
- testing the delivery of an intervention to an individual might raise concerns if others are not included;
- individually-delivered interventions might spread among family, friends or the wider community;

- they are interested in the mass effectiveness of interventions that will be rolled out to the public with varying degrees of quality, uptake, adherence and response, even if they have been shown to be efficacious in individual cases.

These four reasons are important because schools situated in the same town or village have good relationships and communications between each other which can result in “contamination” between the groups. For example, if a school in the control group acquired some information deployed in the intervention group as a result of chatting among parents and school friends, this can motivate control students and families to adopt the intervention recommendations and, thus, they do not behave as controls should, and adversely affect the design of the study.

References:

- National Institute for Health and Care Excellence (NICE). Accessed by: 21st of August 2014 <http://www.nice.org.uk/glossary?letter=R>
- Murray DM. Statistical models appropriate for designs often used in group-randomized trials. *Stat Med* 2001; 20: 1373-85 doi: 10.1002/sim.675 pmid: 11343359.
- Osrin D, Kishwar A, Fernandez A, et al. Ethical challenges in cluster randomized controlled trials: experiences from public health interventions in Africa and Asia. *Bulletin of the World Health Organization* 2009;87:772-779. doi: 10.2471/BLT.08.051060.

Point 2: Inclusion criteria includes name, gender and DOB. Are those missing that information excluded or were some people excluded based on these criteria? If so, why were those with missing names excluded? It doesn't seem as though it would affect the analysis.

Response 2: In Figure 1 of the manuscript we presented the exclusions prior to randomisation and, subsequently, exclusions in the follow-up. These included missing data such as complete (unambiguous) names necessary not only for ethical permission from the ethics committee, but also informed consent from the parent or guardian, and for accurate follow-up.

Point 3: Page 3, line 24. Odds ratios cannot be negative (a similar error is found in the results). Please recheck your statistical results.

Response 3: We apologise for the error. The values we had reported were the coefficients of the models, and not the odds ratios. This has been corrected throughout the revised m/s and, as well, the reported values now correspond to the actual odds ratios.

We have introduced a clarifying statement in Abstract, Pg 3, Ln 23: “. Multivariate statistical analysis indicated that the performance of ≥ 4 after-school PA h/week was a protective factor against childhood OB (OR:0.600; $p=0.032$). ” In the Results section Pg 25, Ln 327 we have added: “. Thus, breakfast dairy product consumption (OR: 0.336; $p=0.004$) and ≥ 4 after-school PA h/week (OR: 0.600; $p=0.032$) were protective factors against OB. Conversely, doing < 4 h/week PA (OR: 1.811; $p=0.018$) increased the risk of childhood OB.

Point 4: Some abbreviations are not defined when they first appear in the text, for example, OB in the abstract

Response 4: We have rechecked all the abbreviations in the manuscript to ensure that no “non-standard” abbreviations appear without prior definition.

Point 5: Page 5, line 60. Which Korea is being mentioned here? I assume it's South Korea.

Response 5: We had used the Obesity Update document published in 2012 by Organization for Economic Co-operation and Development (OCDE) which mentioned the 5 OECD countries where the

obesity epidemic was stable. One was Korea, without clarification whether North or South. However, since North Korea is not a member of the OECD, the inference is that South Korea is the target country. We have clarified this ambiguity by adding "South" Pg 5, Ln 48

References:

1. The Organisation for Economic Co-operation and Development (OECD):Obesity Update 2012. [<http://www.oecd.org/health/49716427.pdf>]

Point 6: Secondary schools are listed in the abstract, yet it appears only primary schools were chosen (page 5, line 83). Please clarify.

Response 6: The present study involved primary schools as described (Pg 7, Ln 116). The schools in Amposta were designated as Group A (intervention) and 9 towns around Amposta (Sant Jaume d'Enveja, Els Muntells, l'Ametlla de Mar, El Perelló, l'Ampolla, Deltebre, l'Aldea, Lligalló del Gànguil and Camarles) as Group B (control).

We deleted the word "secondary" in the abstract Pg 3, Ln 6, and have added: "Setting: primary schools". In Spain, children continue at the same site from the age of 8 until the age of 13 when they proceed to "high school" or Instituto where they continue on to 16 years of age when they leave obligatory schooling and proceed to technical training or to academic courses in preparation for university

Point 7: Page 8, line 166. It's more correct to say the analysis was 'stratified' by gender and not segregated.

Response 7: OK. We have changed this word in the manuscript in Pg 10, Ln 172 and 175 "overall as well as stratified by gender" and in Pg 16, Ln 266: "as well as stratified with respect to gender".

Point 8: Page 9, line 169. How are BMI z-scores computed? Are they based on summary statistics for the overall sample, stratified by gender, or using population values?

Response 8: BMI z-score was computed using the population values of World Health Organization (WHO) obtained from the WHO Global Infobase as relating to children of the same age as our study population. To clarify this point we have added on Pg 10, Ln 180: "BMI z-score was analysed using the population values of the WHO Global InfoBase,[25]."

Point 9: Please provide more information about computing sample size. There are over 60 methods for sample size comparing two proportions.

Response 9: Sample size was computed as customary based on the normal approximation to the binomial distribution.

Point 10: For the linear mixed models, was the outcome proportions of obese? This would assume the proportions are normally distributed. Is that reasonable here?

Response 10: We used the generalised linear mixed model approach with the proportions of obese as outcome in order to avoid the assumption of normal distribution which, in the current case, would be inaccurate, as stated by the Reviewer.

Point 11: It is unclear to me how McNemar's test is used on behavioural changes. Were these values dichotomous and, if so, how?

Response 11: The values assigned to eating habits were dichotomised. We used the Krece Plus questionnaire developed by Serra Majem and colleagues (21). This questionnaire is composed of 15 questions with 2 possible answers: yes or no. The Llargues et al questionnaire regarding lifestyles has 3 possible answers: never, sometimes, always. We only analysed the "never" and "always" answers, while "sometimes" was too ambiguous for our purposes. To clarify this point we have introduced a new paragraph in Methods, Study Population, Pg 8, Ln 135: ". A questionnaire on eating habits (Krece Plus) developed by Serra Majem et al,[21], and physical activity, level of parental education and lifestyles developed by Llargues et al,[22] were filled-in by the parents at baseline and at the end of the study."

References:

21. Serra Majem L, Ribas Barba L, Aranceta Bartrina J, et al. Childhood and adolescent obesity in Spain. Results of EnKid study (1998-2000). *Med Clin (Barc)* 2003;12:725-732.
22. Llargues E, Franco R, Recasens A, et al. Assessment of a school-based intervention in eating habits and physical activity in school children: the AVall study. *J Epidemiol Community Health* 2011;65(10):896-901. doi:10.1136/jech.2009.102319. [Published Online First: 12 March 2011]

Point 12: The authors seem to be assuming data is missing at random. This is not the same as an "implication" or "understood to be".

Response 12: With the type of the intervention used, the drop-outs were not caused by adverse effects of the intervention. Therefore, the statistical methods used are robust, under the assumption of "missingness" at random.

We have changed in Pg 11, Ln 210: "... the assumption being that missing data were random." and in Results section Pg 15, Ln 242: "Drop-outs in both groups are assumed to be missing at random".

Point 13: Tables 2-4. It is unclear to me what is being tested here. Specifically, how is McNemar's test and Fisher's exact test being used?

Response 13: Table 2 compares the changes in obesity and overweight proportions between baseline and end-of-study within each study group (p-value1, McNemar's test) and between groups (p-value2, Fisher's exact test). Table 4 has the same structure but, for clarification, we have named p-value1 the p-value corresponding to the comparison between baseline and end-of-study values in the intervention group; p-value2 is the p-value corresponding to the comparison between baseline and end-of-study values in the control group; p-value3 is the p-value corresponding to the comparison between groups. In Table 3, the structure is the same but there was an error in the footnote (p-value2 is not from Fisher's exact test as originally stated, but from an ANOVA model in which the outcome of the BMI z-score is continuous). The error has been corrected in the revised Table 3.