

## PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Paediatrics Open. The paper was subsequently accepted for publication at BMJ Paediatrics Open.

## ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Effect of the Fukushima earthquake on weight in early childhood - a retrospective analysis
<b>AUTHORS</b>	Ono, Atsushi; Isojima, Tsuyoshi; Yokoya, Susumu; Kato, Noriko; Tanaka, Toshiaki; Yamagata, Zentaro; Chida, Shoichi; Matsubara, Hiroko; Tanaka, Soichiro; Ishikuro, Mami; Kikuya, Masahiro; Kuriyama, Shinichi; Kure, Shigeo; Hosoya, Mitsuki

## VERSION 1 - REVIEW

<b>REVIEWER</b>	Koutoumanou, Eirini University College London, UK Competing interests: None
<b>REVIEW RETURNED</b>	30-Nov-2017

<b>GENERAL COMMENTS</b>	<p>Due to the response bias involved in this study by receiving data from "31 of 57 local governments", I recommend that the authors suggest ways of improving the response rate for future studies. What have they learnt from the experience of collected data from this study that could help researchers in the future?</p> <p>Figure 3 should be presented and explained in the text before figure 2 as otherwise the reader is left with questions as to what groups 0, etc stand for,</p> <p>Please correct the number of the control group to 3206 (shown in text and in figure 2) as opposed to 3200 in figure 3, or vice versa of 3200 is right.</p> <p>I would recommend that the authors use an alternative phrase to "Matched Data" as this implies paired matching with a control/other group. For example use something like "most complete data".</p> <p>The authors should repeat their analysis with data from groups 0 and III included as the following statement is not statistically justified: '...as both Group Zero and Group III include too few children to be used statistically, we excluded the data for two groups as "Unavailable Data" ' - what calculations/assumptions did the authors base this on? No sample size calculations have been presented, hence seems unjustified.</p> <p>Is Table 1 one simply showing the number of boys and girls per group and areas? Make this clear on the table title, explain clearly what the numbers in the brackets are and space the table out so it's so not cramped hence easier to read.</p>
-------------------------	--

	<p>As it is not mentioned, I suspect that the authors have not applied a correction for multiple comparisons when performing the t-tests? Therefore, I strongly recommend that all tests are repeated to ensure that the results have been adjusted for the fact that so many test have been performed (which increases the chance of finding something statistically significant when it is not).</p> <p>The authors mention at the first sentence of the Results section that comparisons are being made before and after the earthquake. In order to be more precise, I recommend that the authors edit this phrase to reflect the fact that comparisons were made between children whose birth/medical assessment was before or after the earthquake.</p> <p>This is the way the authors have introduced/explained their comparison in the earlier sections and it's confusing to alter the way they refer to the comparison at this stage.</p> <p>Did the authors consider fitting a multilevel model accounting for the clustered data within areas and the repeated measures per child? Such model would also allow them to adjust the results for potential confounders. It would have also been interesting to see the trajectory of a group of children from birth to their future measurements before 36 months to get a feel of general trends.</p> <p>Could you please ensure the entire report is proof read by a native English speaker?</p>
--	---

<b>REVIEWER</b>	Tashiro, Satoshi Research Institute for Radiation Biology and Medicine Hiroshima University Japan Competing interests: non
<b>REVIEW RETURNED</b>	11-Dec-2017

<b>GENERAL COMMENTS</b>	<p>It has been suggested that disasters, such as huge earthquake and typhoon, affect children's physical and mental health. In Japan, the Great East Japan Earthquake caused tsunami, which destroyed many houses on the coast and led to the Fukushima Daiichi Nuclear Power Plant Accident. There have been some reports concerning the health effects of the earthquake on adults and children. However, the health effect of the earthquake on early childhood, younger than 3 years old, in Fukushima prefecture is unclear. In this study, the authors examined the Height SDS, BMI-SDS and delta BMI SDS of children in Fukushima prefecture using the data of health examinations. The authors found that overweight was observed between the 18 months and 36-42 months health examinations among the children who experienced the disaster between their birth and 6-10 months health examinations. This trend was observed in Hamadori and Nakadori areas in Fukushima prefecture where the radiation levels were relatively high. This is an important report to show the overweight of children younger than 3 years old in Fukushima, especially in areas with relatively high radiation levels, after the Great East Japan Earthquake. However, authors have to provide more data for better understanding of readers. The reorganization of "introduction" and "discussion" is also required before publication.</p> <p>1) The definition of each group is not clear and confusing. The</p>
-------------------------	---

	<p>authors should provide readers with more data to characterize each group, at least the average and range of ages at the disaster, for better understanding. The Grouping according to ages at the disaster is preferable and will avoid misunderstanding of readers.</p> <p>2) The authors should explain the meaning of “SDS” in the introduction and method sections. How to calculate SDS should be provided in the method section. What is the meaning of “0” in SDS? Without the explanation of SDS, we cannot interpret the results of CI in Tables.</p> <p>3) The authors should provide raw data of Height, Weight and BMI. That would be helpful for readers to understand the data rather than figures only with SDS are provided.</p> <p>4) Most of the 1st paragraph in “discussion,” where why the authors examine BMI in this study is explained, should be moved to “introduction.” The authors can also explain the reason why they use BMI SDS instead of BMI there.</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

#### Reviewer 1

Reviewer’s comments and answers to Reviewer’s comments follow:

#### Comment 1

Due to the response bias involved in this study by receiving data from “31 of 57 local governments”, I recommend that the authors suggest ways of improving the response rate for future studies. What have they learnt from the experience of collected data from this study that could help researchers in the future?

#### Answer 1

The health examination data from within Fukushima prefecture is large, but not all could be obtained and the only common variables available were height and weight. The use of an online backup service and standardization of the format of data sheets to manage health examination records nationwide will likely enable easier collection of more data.

We think that collecting and analyzing baseline data during health examination will be useful for evaluating the effects of an event.

We added the follow sentence to the end of the Discussion section (p.25, l.286–292).

“We retrospectively analyzed the data on health examination and discovered that the availability of normal baseline data for comparison with those after the event were useful. The health examination data from within Fukushima prefecture is large, but not all could be obtained and the only common variables available were height and weight. The use of an online backup service and standardization of the format of data sheets to manage health examination records nationwide will likely enable easier collection of more data.”

#### Comment 2

Figure 3 should be presented and explained in the text before figure 2 as otherwise the reader is left with questions as to what groups 0, etc stand for.

#### Answer 2

As you pointed out, we explained defined the inclusions in the group in the Methods section (p.8, l.118–126). In addition, we reversed the order of Figure 2 and Figure 3 and modified the contents slightly (p.9, l.128–129).

#### Comment 3

Please correct the number of the control group to 3206 (shown in text and in figure 2) as opposed to 3200 in figure 3, or vice versa of 3200 is right.

#### Answer 3

In Figure 2, we corrected the number of the control group from 3,200 to 3,206. In Table 1, we corrected the number of boys in group IV in the Hamadori area from 963 to 1,107.

#### Comment 4

I would recommend that the authors use an alternative phrase to “Matched Data” as this implies paired matching with a control/other group. For example use something like “most complete data”.

#### Answer 4

We changed the phrases “matched data” to “complete data” and “unmatched data” to “incomplete data” (Figure 3).

#### Comment 5

The authors should repeat their analysis with data from groups 0 and III included as the following statement is not statistically justified: ‘...as both Group Zero and Group III include too few children to be used statistically, we excluded the data for two groups as “Unavailable Data” ’ - what calculations/assumptions did the authors base this on? No sample size calculations have been presented, hence seems unjustified.

#### Answer 5

As you pointed out, the reasons for excluding patients were insufficient.

First, group zero represented children who developed the disease before birth; therefore, group zero was not eligible in this survey and was excluded.

A total of 6,167 children who were born between June 1, 2010 and April 30, 2011 were under 10 months old when they experienced the disaster or were not yet born, and most of them had not received their 6th to 10th month health examination at that time. Moreover, 10,046 children who were born between March 1, 2009 and August 31, 2009 were 19–24 months old when they experienced

the disaster, and most of them had already received their 18th month health examination at that time. Therefore, the number of children who satisfied the definition of group III during the first data collection was very few. Because of this small number and the large difference in the populations, group III was not comparable with the other groups under equal conditions and was excluded.

We modified the Study design and subjects part of the Methods section and added the following sentence to that section and Limitations section of the Discussion.

Methods section (Study design and subjects: p.9, l.131–136)

We changed from “Group Zero represents the group of children who were born after the disaster. Moreover, as both Group Zero and Group III include too few children to be used statistically, we excluded the data for two groups as “unavailable data”.” to:

“Group zero represented the group of children who experienced the disease before birth and were, therefore, not eligible and excluded from this survey. The number of children who satisfied the definition of group III during the first data collection was very few. Because of this small number and the large difference in the populations, group III was not comparable with the other groups under equal conditions and was excluded.”

Discussion section (Limitation: p.25, l.281–285)

“Lastly, because the interval between data collections was more than six months, the number of children included in group III became much smaller than the other groups. Consecutive data collection could have minimized the difference between sample sizes of the groups analyzed.”

#### Comment 6

Is Table 1 one simply showing the number of boys and girls per group and areas? Make this clear on the table title, explain clearly what the numbers in the brackets are and space the table out so it's so not cramped hence easier to read.

#### Answer 6

We modified the title of Table 1 to “The sex distribution in each area and group in those with complete data” and explained more clearly the number in the brackets in the Figure Legend.

The title of Table 1 was changed from

“The Matched Data for the six groups”

to

“The sex distribution in each area and group in those with complete data”

#### Footnote of Table 1

※2 the number in brackets represents the number of children who were examined at birth. They are fewer than the number of other health examination periods because some data on height at birth were missing in Nakadori area.

#### Comment 7

As it is not mentioned, I suspect that the authors have not applied a correction for multiple comparisons when performing the t-tests? Therefore, I strongly recommend that all tests are repeated to ensure that the results have been adjusted for the fact that so many test have been performed (which increases the chance of finding something statistically significant when it is not).

#### Answer 7

As you pointed out, the significance level needs to be adjusted because the family-wise error rate increases with multiple comparisons.

In our analysis, we compared the control group with the other three groups (groups I, II, and IV). Because there were three pairwise comparisons, a p value of less than 0.016 was regarded as statistically significant using the Bonferroni correction.

We corrected the Methods section (statistical analysis: p.13, l.174–177) as follows:

From “P value of less than 0.05 was regarded as statistically significant.” to

“In our analysis, because there were three pairwise comparisons between the control group and the other three groups for each analysis stratified by region and sex, a p value of less than 0.016 was regarded as statistically significant using the Bonferroni correction.”

#### Comment 8

The authors mention at the first sentence of the Results section that comparisons are being made before and after the earthquake. In order to be more precise, I recommend that the authors edit this phrase to reflect the fact that comparisons were made between children whose birth/medical assessment was before or after the earthquake.

This is the way the authors have introduced/explained their comparison in the earlier sections and it's confusing to alter the way they refer to the comparison at this stage.

#### Answer 8

As you pointed out, the description of the comparisons was wrong. According to the comment, we modified the Results section (p.14, l.189–192) as follows:

From

“We compared the Height SDS, the BMI SDS and  $\Delta$ BMI SDS between the affected children (Group I, II and IV) and the unaffected children (Control Group) before and after the disaster. Moreover, we compared  $\Delta$ BMI SDS between these groups according to the 3 areas (the Hamadori, Nakadori and Aizu areas).”

To

“We compared the height SDS, BMI SDS, and  $\Delta$ BMI SDS before and after the disaster between the affected children (group I, II and IV) and the unaffected children (control group). Moreover, we compared the  $\Delta$ BMI SDS in these groups according to the three areas (i.e., Hamadori, Nakadori, and Aizu).”

#### Comment 9

Did the authors consider fitting a multilevel model accounting for the clustered data within areas and the repeated measures per child? Such model would also allow them to adjust the results for potential confounders. It would have also been interesting to see the trajectory of a group of children from birth to their future measurements before 36 months to get a feel of general trends.

#### Answer 9

As you pointed out, a multilevel analysis would have enabled us to adjust for multiple factors at different levels. However, our primary objective was to show descriptive data stratified by area and sex; this was clarified in the revised version.

We added the following sentence to the Methods section (Statistical analysis: p.12, l.173–p.13, l.174).

“We stratified the indicators by region and sex in order to examine the differences between the control group and groups I, II, and IV.”

#### Comment 10

Could you please ensure the entire report is proof read by a native English speaker?

#### Answer 10

We ensured that this report was proofread by a native English speaker and attached the certification.

### **Reviewer 2**

Reviewer’s comments and answers to Reviewer’s comments follow:

#### Comment 1

The definition of each group is not clear and confusing. The authors should provide readers with more data to characterize each group, at least the average and range of ages at the disaster, for better understanding. The Grouping according to ages at the disaster is preferable and will avoid misunderstanding of readers.

#### Answer 1

In the Abstract and Methods sections, we clarified the definitions of each group, as follows:

In the Abstract section (p.4, l.52–61), from:

“We divided the affected children into groups according to age at the time of the disaster (Group I, II, III and IV represent children who experienced the disaster between birth and their 3-4 month health examination, their 3-4 and 6-10 month health examinations, their 6-10 and 18 month health examinations and their 18 and 36-42 month health examinations, respectively) and compared the health status of each group with that of “unaffected” controls (children who experienced the disaster after their 36-42 month health examination) using height standard deviation score (Height SDS), body

mass index SDS (BMI SDS) and BMI SDS between two consecutive health examinations ( $\Delta$ BMI SDS).”

To

“We divided the affected children into respective groups according to the interval from the disaster to the time of health examination and age, as follows: group I, birth to 3–4 months in boys [1.81 (range, 0–6 months)] and girls [1.79 (range, 0–7 months)]; group II, 3–4 months to 6–10 months in boys [6.37 (range, 3–9 months)] and girls [6.35 (range, 3–9 months)]; group III, 6–10 months and 18 months in boys [16.2 (range, 5–22 months)] and girls [16.9 (range, 5–22 months)]; and group IV, 18 months to 36–42 months in boys [21.0 (range, 18–24 months)] and girls [21.0 (range, 18–24 months)]. Using height and body mass index, the health status of each group was compared with that of unaffected controls (i.e., children who experienced the disaster after their health examination at 36–42 months).”

In the Methods section (p.8, l.118–p.9, l.128), from

“Moreover, we divided the “Matched Data” into 6 Groups (Figure 3): Group Zero; 551 children who were born after the disaster, Group I; 2114 children who experienced the disaster between birth and their 3-4 month health examinations, Group II; 2888 children who experienced the disaster between their 3-4 and 6-10 month health examinations, Group III; 383 children who experienced the disaster between their 6-10 month and 18 month health examinations, Group IV; 3800 children who experienced the disaster between their 18 month and 36-42 month health examinations and the Control Group; 3206 children who experienced the disaster after their 36-42 month health examination, in order to make comparisons based on the age at which the children had experienced the disaster and their physical activity level.”

To

“We classified the children into the following six groups: 1) group zero included children who experienced the disaster before birth; 2) group I included children who experienced the disaster between birth and their 3–4-month health examination; 3) group II included children who experienced the disaster between their 3–4- and 6–10-month health examinations; 4) group III included children who experienced the disaster between their 6–10-month and 18-month health examinations; 5) group IV included children who experienced the disaster between their 18-month and 36–42-month health examinations; and 6) the control group, which included 3,206 children who experienced the disaster after their 36–42-month health examinations.

We aimed to compare the groups based on the age when the children experienced the disaster and their physical activity level (Figure 2).”

#### Comment 2

The authors should explain the meaning of “SDS” in the introduction and method sections. How to calculate SDS should be provided in the method section. What is the meaning of “0” in SDS? Without the explanation of SDS, we cannot interpret the results of CI in Tables.

#### Answer 2

Standard deviation (SD) is defined as the square root of the variance and represents the width of the distribution. The zero of SD-score (SDS) represents the average value, and the SDS represents the number of times that the SD diverges from the average value.

We modified the following sentence of the Methods section (Statistical analysis: p.12, l.158–170).

From

“We used the height standard deviation score (Height SDS), body mass index SDS (BMI SDS) and BMI SDS between two consecutive health examinations; birth and 3-4 months old, 3-4 and 6-10 months old, and 18 months and 36-42 months old ( $\Delta$ BMI SDS) for statistical analysis. Height SDS and BMI SDS were calculated using calculation software,<sup>12</sup> and  $\Delta$ BMI SDS was generated from the BMI SDS values.”

To

“Due to the fact that BMI in childhood changes substantially with age, comparison of BMI among children of different age groups is difficult. Moreover, the age in months of children taking the same health examination varies. For that reason, it was necessary to standardize each BMI in childhood according to sex and age. Cole et al constructed centile curves for BMI using the LMS method, which was adopted by Inokuchi et al for the Japanese population.[12, 13] Therefore, we are now able to express BMI as SDS. Standard deviation (SD) is defined as the square root of the variance and represents the width of the distribution. The zero of SDS represents the average value, and the SDS represents the number of times that the SD diverges from the average value. We converted the height and BMI of all children to height SDS and BMI SDS using a calculation software,[14] then calculated the  $\Delta$ BMI SDS as the difference in BMI SDS between two consecutive health examinations, as follows: between birth and 3–4 months, between 3–4 and 6–10 months, and between 18 and 36–42 months.”

Comment 3

The authors should provide raw data of Height, Weight and BMI. That would be helpful for readers to understand the data rather than figures only with SDS are provided.

Answer 3

We uploaded the raw data of “height”, “weight”, and “BMI SDS” as a supplemental file.

Comment 4

Most of the 1st paragraph in “discussion,” where why the authors examine BMI in this study is explained, should be moved to “introduction.” The authors can also explain the reason why they use BMI SDS instead of BMI there.

Answer 4

We added the following sentence to the end of the Introduction section (p.6, l.84–88).

“Generally, body mass index (BMI) is used to assess physical data, such as obesity, in adults. However, BMI cannot be compared among children because it varies in different age groups and changes substantially with age. In order to standardize the height and BMI in childhood according to sex and age, we used the standard deviation score (SDS).”

The other clarifications about SDS were explained in our answer to your comment #2.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Koutoumanou, Eirini UCL, London Competing interests: none
<b>REVIEW RETURNED</b>	11-Jan-2018

<b>GENERAL COMMENTS</b>	The authors have done a great job editing their manuscript in response to the reviewers' comments. I believe they have produced a better manuscript which is ready for publication.
-------------------------	---