PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The cost of mental and physical health disability in childhood and adolescence to families in the United Kingdom: findings from a repeated cross-sectional survey using propensity score matching
AUTHORS	Solmi, Francesca; Melnychuk, Mariya; Morris, Stephen

VERSION 1 – REVIEW

REVIEWER	Luke Munford Manchester Centre for Health Economics, University of Manchester, UK
REVIEW RETURNED	03-Aug-2017

GENERAL COMMENTS	Thank you for the chance to review the article titled "The cost of mental health disability to families in childhood and adolescence in the United Kingdom: findings from a repeated cross-sectional survey using propensity score matching". This is a very interesting, well written article. The study elicits a value for the compensating variation (CV) a family with a child with mental health disability(ies) would require in order for them to achieve similar living standards to a ceteris paribus identical family with a non-mentally disabled child. The authors consider different definitions of mental disability to allow for comorbidities and different combinations of the variables use to define mental disability (two out of eight self-reported disabilities are defined as mental health related). This leads to many comparisons. The authors come up with different values of the CV depending on which group is being compared. They then compare these values to actual benefit receipts, for the whole sample and then split by living-standards (high vs. low). This comparison leads to the conclusion that means testing benefits could be more equitable. I have a few points on the study, listed below. Points for consideration: On pages 8 and 9 (lines 234 – 238) you infer that you include benefits received in your calculation of income. However, in the 'Benefits' subsection of the 'Results' section (page 12; lines 349 – 361) you seem to infer that (in some cases) the benefits closely match the extra income needed. But if you have already accounted for benefits in your income calculation, does this not then imply that benefits should be/need to be doubled? My understanding is conditional on benefits received, at least the same again is needed to remove the CV. You discuss this point on

	lines 427 – 432, but I am not sure I follow the reasoning. Is the point that living standards are a function of actual, reported, income, which is itself a function of benefits received? I think this impacts on the interpretation of your results. You need to be careful you do not double count benefits and assert they are equal to the CV, when in fact they're already accounted for, so perhaps are not sufficient. On line 436 you state that your propensity score model could have been improved by including (inter alia) parental employment status. Is there a reason why you did not do this? If it is in the data, then I think you should. If one (or possibly both) parent(s) cannot work because of the disability of the child, then I think this will affect your results. Comparing incomes without accounting for employment status seems a little strange to me; is employment (or income form employment) not the major contributor to overall income? Your conclusion that benefits should be means tested is only based on a few of your comparisons (Table 3; bold results), not all. So perhaps this needs stressing more. If you are basing your conclusions only on these results (and ignoring the statistically insignificant ones), I think you should justify why more. Minor points: Add confidence intervals to the point estimates in the abstract. Line 60: either "…help achieve a more…" or "…help in achieving a more…" Line 63: an 'a' is missing ("…(i): a large…") Line 104: 'y' missing from "risky" Line 174: group 4 I think it should be plural, i.e. "Both mental disabilities…" On lines 261 – 262 you state that this method has been used in adults. However, one of your earlier references (6) uses this method in children too. Perhaps you should mention that here. If my understanding is correct, you create your matched samples, and then run the following regression: $⟦LSI⟧_i=αT_i+βY_i+δ⟦(T*Y)⟧_i+γX_i+ε_i$ where T is a treatment indicator (=1 if child meets criteria; =0 if non-disabled, say), Y is income, and X is a vector of other characteristics. Then the CV is the parameter δ? If this is the case, shouldn't the sample size by twice that reported in Table 3 (i.e. currently you report the number of people in each group (as in table 1), but shouldn't this be then double if you match each treated family to a matched control?) It would be good to discuss non-statistically significant results a little more. These can be as interesting as the significant ones. For example, from table 3 it would appear that people with children with only mental disabilities are suitably compensated (insignificant effect for comparison A). The sign of the CV is actually negative, although

	with a huge confidence interval. On line 336 you mention groups 6 – 8, but in the earlier part of the paper only six groups are introduced. What are groups 7 and 8? Table 1 is a little unclear when it comes to the p-values. For example, in column (3) – group 1 – what does p<0.0001 on the 'Total' line mean? I assume it refers to the information below, that there are gender differences? But this is unclear. Also, what are you testing to what? This could do with further explanation – is the value above comparting group 1 to group 6? Or group 1 to all? This information could be added in a Table note. Also on table 1, the way education is displayed is not clear. Would it not be easier to add 18 to everything? Or change the label? At present it is age, but I assume this is years (post-18). Table 2 – it may be more clear if in the square brackets you put "[N=##]" instead of "[##]". Also, in the Table note, could you elaborate and say something like "* indicates significantly different from reference group (Group 6) at p<=0.05". As it currently reads it is not clear that this is what you mean. In Table 4, if you repeat this for Group 6 (reference group), is the mean value of benefits received zero, as expected? Or not? This need not be reported in the paper, but would be interesting to see if the comparison group is picking up benefits received from elsewhere. In rows (1), (2), (3), and (5) what do the median values =0 mean? That the median value of benefit received is zero. If so, does this perhaps need some discussion? Why would the median be zero? That means over 50% of families in this group receive zero benefits? Is that plausible? Supplementary file 1 – CV: "…(i) the curves are upward sloping from left to right, due to diminishing returns to S as Y increases" This needs adding to. Simply being upward sloping is not sufficient to ensure diminishing returns. You need to mention that the curves are convex (a concave function is upward sloping but exhibits increasing returns). Supplementary file 2 – PSM: do you use 1:1 nearest neighbour matching? If so, could you add this, for completeness?

REVIEWER	William Whittaker University of Manchester, UK
REVIEW RETURNED	16-Aug-2017

GENERAL COMMENTS	The study seeks to estimate the cost of child mental disability. To do this the study uses data from eight waves of the Family Resources Survey, a repeated cross-sectional survey. Families with and without a disabled child are matched via propensity score matching techniques, with the aim that any variation in income represents the additional income required to compensate for the presence of a disabled child. Additional analyses compared compensating variation by sub-groups of deprivation. On average, £33 compensation was required to match the living standards of those with a mental health disabled child to those without, this varied across deprivation groups and with comorbidity. £25 compensation variation was calculated for deprived families with a physically disabled child to those without. The authors conclude that deprived families have been undercompensated under the current benefit system and that means testing may achieve a more equitable redistribution of benefits. Overall summary The study's introduction highlights the importance of the topic, a high burden of illness that has been growing and persists into adulthood. A clear argument is made for focussing on inequalities too. The institutional set up is clear. The research question is an important one and the study makes a significant contribution to the literature, being the first to be able to measure these compensation variations by mental and physical disability status. There are clear policy implications of the study and overall the authors have taken a comprehensive and thorough approach to identify causation. Overall I think this is a very good manuscript. There are three areas in which I feel the manuscript could be improved, these are largely for clarification purposes. Unit of analysis: At times it was difficult to understand the structure of the data and unit of analysis. • Materials and Methods: Page 5: could the authors state the volume of households in the FRS (approximately 20,000 per year) and clarify the unit of analysis throughout? There is a jump from households to families to children that isn't clearly detailed in Page 5 line 51 to Page 6 line 15. Maybe explaining the volume of families and individuals and children in the FRS may be sufficient in clarifying this. It is in the Results section that the reader realises this is child level unit of analysis so anywhere this can be made more explicit in the data section would help. In addition, is there potential for multiple children in the same family to be in the dataset and have the authors clustered for this? • Results: Page 10, line 32: The authors state the initial sample concerns 99,142 children, but in line with my above request for clarity, it would be helpful to give the volume of families too. Propensity Score Matching: In general I felt more detail and checks of robustness and validity were needed. • Page 10, line 3 to line 18: the authors state earlier (Page 8, line

	46) that the Living Standards Index (LSI) was used in the PSM models but it is not mentioned here. In the Supplementary Material I found it difficult to understand what the authors meant by: “we included values of our LSI rounded to the first decimal point as external to X and matched on both” (Page 25, line 25), in particular what do the authors mean by ‘as external to X and matched on both’? Later it appears LSI was used to stratify analysis (Page 11, line 37), is this what the authors mean? If so, why didn’t the authors match on LSI? • Supplementary material: Page 25: There are alternative PSM techniques (many to one; radius; kernel), additional sensitivity analysis could have investigated this. In addition, the results of the matching (e.g. comparison of means and particularly variance) would be insightful and reassure the reader. Benefits approach/policy implications: Clarification is needed as to the implications of income including child disability benefit. • Page 8, line 52 to Page 9 line 11: Does the income measure include child disability benefits? If disability benefits offset the additional cost of disability then no contingency variation should be found. Since disability benefits are included in the income measure the authors findings of contingency variation are thus over and above those benefits currently claimed. If my interpretation is correct (it may not be!) then the comparison to the amount of disability benefits received discussed in the Methods on Page 10 line 22 and Results on Page 12 lines 21 to 40, hold little value. Note however that should this interpretation be valid then the results are all the more important since they suggest an even greater need for social assistance than is currently being argued in the study. Minor clarifications: • Page 5, lines 9-11: It might strengthen the clarity of the policy relevance of the study if the authors were to better link the discussion of Disability Living Allowance (DLA) to the compensating variation estimated. Are the authors seeking to question the appropriateness of the DLA levels and/or to see whether an additional stratification of benefit based on mental or physical disability is needed? • Page 6, line 52 to Page 8 line 5: The groupings of children appears sensible, if somewhat difficult to follow. I drew a Venn diagram to assist my understanding, this involved 4 circles (two mental disabilities, one for physical disability and one for no disability with the three disabilities intersecting). It may be worth considering something like this as a supplementary piece should other reviewers have had similar confusion. • Page 8, line15: Can the authors clarify that selection of the questions comprising the living standards index was based only on data availability and not authors selection of questions deemed relevant? • Page 24, line 25: Figure 1 should be Figure S1 • Table 1: Can the authors state what BU stands for in ‘Adult with disability in BU’? A proportion of the families in adjacent years are followed by the FRS meaning the samples are not independent, whilst there is little the authors can do about this, it could be noted in the study as per the ONS recommendations on Page 13 of: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/321821/uncertainty-family-resources-survey-based-analysis.pdf

VERSION 1 – AUTHOR RESPONSE

Reviewer 1
I have a few points on the study, listed below.
Points for consideration:

• On pages 8 and 9 (lines 234 – 238) you infer that you include benefits received in your calculation of income. However, in the 'Benefits' subsection of the 'Results' section (page 12; lines 349 – 361) you seem to infer that (in some cases) the benefits closely match the extra income needed. But if you have already accounted for benefits in your income calculation, does this not then imply that benefits should be/need to be doubled? My understanding is conditional on benefits received, at least the same again is needed to remove the CV. You discuss this point on lines 427 – 432, but I am not sure I follow the reasoning. Is the point that living standards are a function of actual, reported, income, which is itself a function of benefits received? I think this impacts on the interpretation of your results. You need to be careful you do not double count benefits and assert they are equal to the CV, when in fact they're already accounted for, so perhaps are not sufficient.

Response: The reviewer is correct. In our measure of income we did include benefits, so the compensating variation itself (as opposed to the difference between the compensating variation and the amount of DLA benefits received) should be thought of as the amount by which families are under-compensated under the current benefit system. We have clarified this in the text and removed sections were we previously compared the compensating variation to the amount of benefits received on page 2 (abstract); page 12 (Benefits section), and in the discussion on page 13-16.

• On line 436 you state that your propensity score model could have been improved by including (inter alia) parental employment status. Is there a reason why you did not do this? If it is in the data, then I think you should. If one (or possibly both) parent(s) cannot work because of the disability of the child, then I think this will affect your results. Comparing incomes without accounting for employment status seems a little strange to me; is employment (or income form employment) not the major contributor to overall income?

Response:The reviewer makes an important point for which we are grateful. We have now included parental employment status, which has led to improved matching and resulted in greater income differences. We have included a reference to this variable on page 9 lines 268-269. In our discussion (page 15, lines 462-463) we still suggest that a better indicator of parental physical or mental health could have improved our model.

• Your conclusion that benefits should be means tested is only based on a few of your comparisons (Table 3; bold results), not all. So perhaps this needs stressing more. If you are basing your conclusions only on these results (and ignoring the statistically insignificant ones), I think you should justify why more.

Response:The reviewer is correct, we are considering as 'real' income differences only those whose 95% confidence interval did not include zero and hence had a p-value lower or equal to 0.05 (denoted by ** in Table 4). We have now added 95%CI in the text around these figures to clarify this (page 12-13. Lines 352-382) as well as a sentence explaining that we did not find any other associations (page 13, lines 381-382).

Minor points:
• Add confidence intervals to the point estimates in the abstract.

Response: We have now included these on page 2 lines 49-58.

• Line 60: either "…help achieve a more…" or "…help in achieving a more…"

Response: We have amended the wording to "help achieve a more…" (page 2, line 62)

• Line 63: an 'a' is missing ("…(i): a large…")

Response: We have included this (page 3, line 64)

• Line 104: 'y' missing from "risky"

Response: This has been changed on page 4, line 105.

• Line 174: group 4 I think it should be plural, i.e. "Both mental disabilities…"

Response: We have changed to word disability to disabilities on page 7 line 183

• On lines 261 – 262 you state that this method has been used in adults. However, one of your earlier references (6) uses this method in children too. Perhaps you should mention that here.

Response: We have now included reference 6 to explain that this approach has been used in previous studies of child disability on page 10, line 279.

• If my understanding is correct, you create your matched samples, and then run the following regression: [] where T is a treatment indicator (=1 if child meets criteria; =0 if non-disabled, say), Y is income, and X is a vector of other characteristics. Then the CV is the parameter δ? If this is the case, shouldn't the sample size by twice that reported in Table 3 (i.e. currently you report the number of people in each group (as in table 1), but shouldn't this be then double if you match each treated family to a matched control?)

Response: The reviewer is correct. The numbers reported in square brackets in Table 4 (previously table 3) refer to the number of cases on common support (i.e. included) in the model over the total number of cases (also reported in Table 1). Each case was then matched to a control, so that the actual sample size of each model is twice that on common support. We now have added a footnote to Table 4 explaining this, and changed the column headings to clarify the nature of the numbers in brackets.

• It would be good to discuss non-statistically significant results a little more. These can be as interesting as the significant ones. For example, from table 3 it would appear that people with children with only mental disabilities are suitably compensated (insignificant effect for comparison A). The sign of the CV is actually negative, although with a huge confidence interval.

Response: We have now added a line describing these findings on page 13 lines 377 – 382.

• On line 336 you mention groups 6 – 8, but in the earlier part of the paper only six groups are introduced. What are groups 7 and 8?

Response: We thank the reviewer for noticing this error. The line in brackets was meant to read (i.e. comparisons F – H). We have now amended the manuscript accordingly on page 13 lines 374 – 375.

• Table 1 is a little unclear when it comes to the p-values. For example, in column (3) – group 1 – what does p<0.0001 on the 'Total' line mean? I assume it refers to the information below, that there are gender differences? But this is unclear. Also, what are you testing to what? This could do with further explanation – is the value above comparting group 1 to group 6? Or group 1 to all? This information could be added in a Table note. Also on table 1, the way education is displayed is not clear. Would it not be easier to add 18 to everything? Or change the label? At present it is age, but I assume this is years (post-18).

Response: We have now better formatted Table 1 so that it is clearer that the p-values refer to the variables that have been cross-tabulated between our different disability groups, and have added a footnote to this effect. We have also changed the reporting of our education variable results (adding 18 to each value).

• Table 2 – it may be more clear if in the square brackets you put "[N=##]" instead of "[##]". Also, in the Table note, could you elaborate and say something like "* indicates significantly different from reference group (Group 6) at p<=0.05". As it currently reads it is not clear that this is what you mean.

Response: For consistency with table 4, we have now changed the column headings of table 2 to specify that the value in brackets represents sample size [N]. In the table footnote we additionally specify that "Numbers in squared brackets are sample sizes." Finally, we have clarified the footnote with respect to the p-value ≤ 0.05.

• In Table 4, if you repeat this for Group 6 (reference group), is the mean value of benefits received zero, as expected? Or not? This need not be reported in the paper, but would be interesting to see if the comparison group is picking up benefits received from elsewhere. In rows (1), (2), (3), and (5) what do the median values =0 mean? That the median value of benefit received is zero. If so, does this perhaps need some discussion? Why would the median be zero? That means over 50% of families in this group receive zero benefits? Is that plausible?

Response: The reviewer raises an important point. We believe that it is plausible that over half of the families do not receive benefits because our definitions do not account for severity of disability, which is assessed to decide whether (and in which amount) DLA benefits should be granted. Supporting this hypothesis, median value of benefits received by families of children with both mental health disabilities (arguably a more severely disabled group) is not zero and higher than the mean. No children in our non-disabled group were receiving any disability benefits. We now specify this on page 12 lines 347-348. Please note that now we have re-numbered table 4 as table 3, as we are presenting results relative to benefits earlier in the manuscript.

• Supplementary file 1 – CV: "… (i) the curves are upward sloping from left to right, due to diminishing returns to S as Y increases" This needs adding to. Simply being upward sloping is not sufficient to ensure diminishing returns. You need to mention that the curves are convex (a concave function is upward sloping but exhibits increasing returns).

Response: We thank the reviewer for noticing this. We have now added that the curves are convex on page 3 line 22 of supplementary material.

• Supplementary file 2 – PSM: do you use 1:1 nearest neighbour matching? If so, could you add this, for completeness?

Response: We have added that we are using 1:1 matching on page 4 line 51 of the supplementary material.

REVIEWER 2

Overall I think this is a very good manuscript. There are three areas in which I feel the manuscript could be improved, these are largely for clarification purposes.

Unit of analysis:

At times it was difficult to understand the structure of the data and unit of analysis.

• Materials and Methods: Page 5: could the authors state the volume of households in the FRS (approximately 20,000 per year) and clarify the unit of analysis throughout? There is a jump from households to families to children that isn't clearly detailed in Page 5 line 51 to Page 6 line 15. Maybe explaining the volume of families and individuals and children in the FRS may be sufficient in clarifying this. It is in the Results section that the reader realises this is child level unit of analysis so anywhere this can be made more explicit in the data section would help. In addition, is there potential for multiple children in the same family to be in the dataset and have the authors clustered for this?

Response: We agree with the reviewer that the structure of our sample was not sufficiently clear in our original manuscript. We have now included a paragraph explaining the structure of the FRS on page 6 lines 143-148. Briefly, the FRS collects data on all individuals (i.e. adults and children) living within benefit units (i.e. families), which are, in turn, nested within households. Since in our analyses we match on presence/absence of child disability, we considered children our main unit of analyses, although income data refers to their families. Hence we say that we included children and their families on page 6 line 151.

In terms of family clustering, in our main analyses, each disabled child was matched to only 1 child, hence we did not cluster our analyses by family.

• Results: Page 10, line 32: The authors state the initial sample concerns 99,142 children, but in line with my above request for clarity, it would be helpful to give the volume of families too.

Response: We agree with the reviewer that this information would improve the clarity of the manuscript. On page 11 line 306 we now explain that these children were nested in 61,952 families and, on page 11 lines 310-311, that our complete case sample consisted of "85,212 children nested in 52,639 families (with minimum of 1 and maximum of 8 children)".

Propensity Score Matching:

In general I felt more detail and checks of robustness and validity were needed.

• Page 10, line 3 to line 18: the authors state earlier (Page 8, line 46) that the Living Standards Index (LSI) was used in the PSM models but it is not mentioned here. In the Supplementary Material I found it difficult to understand what the authors meant by: "we included values of our LSI rounded to the first decimal point as external to X and matched on both" (Page 25, line 25), in particular what do the authors mean by 'as external to X and matched on both'? Later it appears LSI was used to stratify analysis (Page 11, line 37), is this what the authors mean? If so, why didn't the authors match on LSI?

Response: We thank the reviewer for this comment which we believe will help clarifying the methodology. We wanted to achieve exact matching on LSI, which is why we included it in the matching algorithm but not in the PSM calculations. The procedure we used was as follows. First, the propensity score was calculated as the predicted probability from the probit model. Then, for each family with a disabled child we selected a match from the pool of families without a disabled child with the same value of the living standards index (based on the first four digits of the index) and the closest propensity score within the common support area. We explain this on page 4 lines 60-63 of the supplemental material.

• Supplementary material: Page 25: There are alternative PSM techniques (many to one; radius; kernel), additional sensitivity analysis could have investigated this. In addition, the results of the matching (e.g. comparison of means and particularly variance) would be insightful and reassure the reader.

Response: We agree with the reviewer that reporting sensitivity analyses would improve our study. We have now calculated the compensating variation using different matching approaches (i.e., 3:1 matching, and radius matching), as described in the methods section pages 10-11 lines 299 - 303. For each matching model, we evaluated whether the groups being compared were balanced; although in some of the models in our main analysis (Comparisons F, G, H) there were some differences in characteristics between our 'treated' and 'untreated' groups, these were not present in sensitivity analyses using radius matching, (whose findings were largely consistent with those of our main analyses). Hence we do not believe that this is likely to have biased our results. We now explain this on page 13 lines 384 – 395.

The results of sensitivity analyses are largely in agreement with those of our main models, and because of the high number of comparisons that we are undertaking in our result sections we only report the findings which differed in sensitivity analyses using radius matching (pages 10-11, lines 300-303).

Benefits approach/policy implications:
Clarification is needed as to the implications of income including child disability benefit.
• •Page 8, line 52 to Page 9 line 11: Does the income measure include child disability benefits? If disability benefits offset the additional cost of disability then no contingency variation should be found. Since disability benefits are included in the income measure the authors findings of contingency variation are thus over and above those benefits currently claimed. If my interpretation is correct (it may not be!) then the comparison to the amount of disability benefits received discussed in the Methods on Page 10 line 22 and Results on Page 12 lines 21 to 40, hold little value. Note however that should this interpretation be valid then the results are all the more important since they suggest an even greater need for social assistance than is currently being argued in the study.

Response: The reviewer is correct, and this is something which was pointed out by reviewer 1 as well. We have now removed all instances in which we compared benefits received to the compensating variation; for more details we refer to our answer to reviewer 1.

Minor clarifications:
• Page 5, lines 9-11: It might strengthen the clarity of the policy relevance of the study if the authors were to better link the discussion of Disability Living Allowance (DLA) to the compensating variation estimated. Are the authors seeking to question the appropriateness of the DLA levels and/or to see whether an additional stratification of benefit based on mental or physical disability is needed?

Response: We have now added a reference to this point, which we agree would help strengthen the argument in support of our paper, in the instruction. On page 5 lines 125-129 we say that we aim to "compare the costs of mental health versus physical health disability to assess whether such stratification is needed" and to "examine how these costs vary by economic deprivation in order to assess whether means testing of DLA should be considered"

• Page 6, line 52 to Page 8 line 5: The groupings of children appears sensible, if somewhat difficult to follow. I drew a Venn diagram to assist my understanding, this involved 4 circles (two mental disabilities, one for physical disability and one for no disability with the three disabilities intersecting). It may be worth considering something like this as a supplementary piece should other reviewers have had similar confusion.

Response: We appreciate that our groups can be complicated to follow. We have now added Table S1 in supplementary material where visually show the composition of our groups. We refer to this on page 7 lines 187 and hope this helps the reader to understand the different groups.

• Page 8, line15: Can the authors clarify that selection of the questions comprising the living standards index was based only on data availability and not authors selection of questions deemed relevant?

Response: The reviewer raises a good point. We now provide greater detail on how these questions were selected on page 8 lines 221 - 228. We also include Table S2 in supplementary material, showing the list of questions included in the FRS and highlighting which ones we employed for our index. For the ones that were not included we provide a rationale for our choice.

• Page 24, line 25: Figure 1 should be Figure S1

Response: We have now changed this (on page 3 of supplemental material)

• Table 1: Can the authors state what BU stands for in 'Adult with disability in BU'?

Response: The abbreviation BU was meant for 'benefit unit'. However, since we refer to benefit units as families throughout we have now changed this to 'Adult with disability in family'

A proportion of the families in adjacent years are followed by the FRS meaning the samples are not independent, whilst there is little the authors can do about this, it could be noted in the study as per the ONS recommendations on Page 13 of:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/321821/uncertainty-family-resources-survey-based

We thank the reviewer for noting this. We have now listed this as a limitation on page 16 lines 472 – 475.

VERSION 2 – REVIEW

REVIEWER	Luke Munford Manchester Centre for Health Economics, University of Manchester, UK
REVIEW RETURNED	23-Oct-2017

GENERAL COMMENTS	Thank you for the chance to review the revised version of the article titled "The cost of mental and physical health disability to families in childhood and adolescence to families in the United Kingdom: findings from a repeated cross-sectional survey using propensity score matching". This is a very interesting, well written article. I wish to thank the author(s) for addressing the points I raised on the initial version. As a result, I believe the paper now reads much better, and the key message is clearer. In particular, I think it is now much clearer that the CV is additional to the benefits already received. And controlling for respondent (parental) employment status makes the comparisons stronger (a point indicated by the author(s) in that their matching is now improved). Finally, the tables are now much more comprehensible and hence easier to interpret.

REVIEWER	Will Whittaker University of Manchester, UK
REVIEW RETURNED	04-Oct-2017

GENERAL COMMENTS	Many thanks to the authors for a very clear guide to the changes made during the revision of the paper. I still consider this to be a very good paper and one with important implications for policy. I am satisfied with the revisions made to clarify the unit of analysis. The additional propensity score matching sensitivity analyses addresses my concern of sensitivity of the results to the matching process taken. It would be informative to give the covariate distributions of each treatment group though I understand this is a tall order requiring 2x3(income stratifications)x8(groups) sets of summary statistics. I agree with the authors that this is unnecessary and the supplementary Table S3 is sufficient enough to reassure readers querying the success of the matching approach. I am satisfied with the clarifications made concerning whether disability benefits were included in the measure of income. I particularly like the use of the terms under-compensated and over-compensated. I have a few very minor suggestions: • Abstract: Results, lines 50 and 51: 'either mental health disability' and 'both mental health disabilities': the variants in mental health disability are not noted in the abstract so this is unclear in isolation of the manuscript. 'either definition of mental health disability' in line 50 might fix this. • Line 287: There is a discrepancy in lines 287 and 293 – A-F are mentioned as matched but in 293 it appears A-H are matched.

	Matching appears to have occurred for A-H so I suspect this is a typo in line 287. • Line 399: ‘analysis we similar’ change to ‘analysis were similar’ • Line 417: ‘disability, was higher’ could be ‘disability, higher’

VERSION 2 – AUTHOR RESPONSE

We have addressed the minor comments raised by reviewer 2, as detailed below.

• Abstract: Results, lines 50 and 51: ‘either mental health disability’ and ‘both mental health disabilities’: the variants in mental health disability are not noted in the abstract so this is unclear in isolation of the manuscript. ‘either definition of mental health disability’ in line 50 might fix this.

Response: The reviewer is correct that it is not clear what the two categories refer to in the abstract. We have now reworded the sentence as follows: "Families of a child with any mental health disability, regardless of the presence of physical health comorbidity, needed an additional £49.31 (95% confidence interval (CI): 21.95; 76.67) and, for more severe disabilities, an additional £57.56 (95%CI: 17.69; 97.44) per week [...]". We hope this improves readability of the abstract.

• Line 287: There is a discrepancy in lines 287 and 293 – A-F are mentioned as matched but in 293 it appears A-H are matched. Matching appears to have occurred for A-H so I suspect this is a typo in line 287.

Response: We have now changed "A-F" into "A-H" and thank the reviewer for noticing this discrepancy.

• Line 399: ‘analysis we similar’ change to ‘analysis were similar’

Response: We have now amended this in the manuscript.

• Line 417: ‘disability, was higher’ could be ‘disability, higher’

Response: We have changed this in the manuscript.

All these changes have been highlighted in yellow in the manuscript for clarity.