Author's response to reviews

Title: Can we trust measures of healthcare utilisation from household surveys?

Authors:

Evelyn K Ansah (Ansahékdr@yahoo.co.uk)
Timothy Powell-Jackson (Timothy.Powell-Jackson@lshtm.ac.uk)

Version: 3 Date: 24 July 2013

Author's response to reviews: see over
20\textsuperscript{th} July, 2013

Dear Editor,

\textit{Re: MS: 1105975019943359- Can we trust measures of healthcare utilization from household surveys?}

We are writing in response to comments provided by the peer reviewers on our paper submitted to BMC Public Health. We greatly appreciate the detailed and insightful comments offered by the reviewers. They have proved very helpful in strengthening the paper. We are positive in all of our responses and in no instance do we offer a rebuttal. Please find our responses to each of the points made by the reviewers outlined below.

\textbf{REVIEWER 1}

1. \textit{A prominent paper that has addressed similar concerns of recall is by Das, Hammer, & Sanchez JDE 2012. The authors mention this paper in the discussion – would be helpful to cover this in the literature summary and also point out what the current paper adds in light of Das et al.}

We now mention this paper in the background section and highlight what our paper adds over and above that the recent literature (page 5, para 1). Our main contribution is to examine whether the method of data collection makes any difference to the results in a free healthcare experiment. The conclusions from the original experiment had real policy implications. Hence the fact that one of the key results does not hold when we use household survey data is an important methodological point.

2. \textit{Figure 3 reports morbidity / illness episodes across time periods from the two data sources. If I am reading this correctly, the graphs on the right side show % of households that reported illness in past 6 months. The panels on the left, however, report number of episodes over a year (diary) and over a month (survey). It would be helpful to standardize the two for comparison – I realize this is not ideal, but would make the discussion on page 10 on pattern of morbidity more transparent.}

We have revised the paper in response to point 7 made by the reviewer. We therefore no longer have any graphs showing morbidity by wealth group. We fully agree with the more general point about having comparable measures across the two data collection methods and have revised the analysis accordingly. Please see our response to point 8 and 9 which provide the full details of how we responded to this comment.

3. \textit{The authors report that the figure 4 shows evidence of a “modest negative wealth gradient” using the diary, and a small positive relationship with wealth. I would recommend that the authors demonstrate some statistical test of significance before one can make this inference. Looking at numbers in the figure, the % of households that report a visit in 6 months in the diary seems very...}
The number of visits per year appears to vary between 2.6 and 2.4 visits per year. Similarly, for household survey, the visits per year appear to vary from just below to just above 0.2. (I assume this is % of households that report any visits in last year?)

We have revised the paper in response to point 7 made by the reviewer. We therefore no longer have any graphs showing morbidity by wealth group.

4. On the same point, the authors mention about a reversal in gradient in the discussion section. Without a significant difference (in magnitude and statistical significance) in levels of utilization, it might be premature to draw conclusions about a gradient or its reversal.

Again, please refer to our response to point 7.

5. The authors also mention that the patterns observed raise question about positive income elasticity of demand for healthcare. I am not sure the paper presents adequate analysis to support this conclusion. Looking only at the data from diary, figure 3 shows a small negative gradient for annual morbidity, and figure 4 shows a relatively flat relationship between visits per year and wealth quintiles (between 2.6 and 2.4). Taken together, wouldn't this suggest the opposite: the rich are less sick, but tend to use about the same level of healthcare.

We agree, please refer to our response to point 7.

6. Table 2 probably has a typographical error, the CI for any clinic visit in Panel A appears to include a negative sign before the 0.006

The original was correct. The point estimate is just significant at the 10% level so the 95% CI crosses zero, hence a lower bound -0.006 and an upper bound 0.075.

7. Overall, I would suggest focusing this paper on recall issues to keep it tight and avoid unnecessary distraction. The figures on wealth quintile morbidity or utilization patterns do not add to the understanding of recall problems in surveys.

We agree the descriptive results showing patterns in morbidity and utilization across wealth groups do not add to the understanding of recall problems. We have removed this analysis from the paper. We now have a tighter focus on how the effect of free care on utilization depends on the data collection method.

REVIEWER 2

8. i) Measures of Morbidity: My main suggestion is clarifying precisely what was done, and what can be understood from the data. The suggestion is motivated, in part, by the difficulty of parsing out the duration of recall and potential seasonal effects from the impact of the data source. Here is what I was thinking: Imagine that you have two time periods, and no seasonal effects, so that morbidity and health seeking are identical in both. The ‘true’ data is that, in a population of 100 families, 20 have an illness only in T1, 20 have an illness only in T2 and 40 have an illness in both T1 and T2. With no biases, the household survey will then pick up the following report: 20 households will report an illness in T1 60 households will report an illness in T2, since this is their most ‘recent’
illness. ii) As far as I understand it, the pictorial diary will actually pick up 120 illness episodes (20 + 20 + 40*2). The difference of 40 comes because the 40 families who had an illness in both T1 and T2 will report only the T2 illness in the household survey. iv) For instance, if T1 is the low sickness period and T2 the high, then the household survey estimates would have to be appropriately weighted by the timing of the survey. Again, the best comparison would be a period-by-period comparison. So, for instance, for households surveyed in T1, I would suggest comparing the number of households who did not seek any care/report any morbidity in the one month preceding using the household survey and the pictorial diary, and similarly for T2. v) In sum, I would much prefer using two specific measures: 1. The percentage of households who did not report any morbidity in the one month preceding the survey. 2. The percentage of households who did not report any morbidity in the three months preceding the survey. The same measures can then be computed for health utilization. To adjust for seasonality, I would pick the appropriate months correctly. For instance, if the pictorial diary was conducted from January to June, and some households were surveyed in May, others in June, then for the May households I would compare the pictorial diary only for May and for March, April and May. I am assuming that no households were surveyed in July (that is, after the final collection of the pictorial diary) since this will constitute missing data from one source.

We agree with this point and have done our best to incorporate it into our revisions. Our response also covers point 2 made by the other reviewer. As a consequence we have made changes to the methods section (page 7-9), the results section (page 10-11) and the discussion (page 13). Specifically, we have done the following. First, we have constructed a new set of variables from the pictorial diary data to allow for a direct comparison with the household survey data. These measures, defined in the same way, are as follows: any illness in the past one month, any clinic visit in the past one month and any informal care visit in the past one month. To address the issue of seasonality, we use data from the November pictorial diary to correspond with the month of recall in the household survey (conducted in December of the same year). Unfortunately we are unable to construct morbidity and utilization measures in the past three months because the data in the household survey only allow for a one week, two week, one month and one year recall. Second, we also produce results for informal care visits. Our hypothesis is that the recall of primary care visits by families in the intervention group deteriorated precisely because they no longer had to pay for healthcare, making the event less salient and more easily forgotten. Informal care visits provide (in a sense) a placebo test of this hypothesis. Because the intervention did not affect the price of care at informal providers the salience and recollection of informal care seeking events should not have been affected by the intervention. And indeed the findings with respect to informal care are consistent with this line of argument.

9. Reporting Results (Table 2): Using these two measures will make Table 2 easier to understand. Currently, Table 2 reports “Clinic Visits per year” and “Any clinic visit” for the pictorial diary and “Any clinic visit past” (which is not properly defined) and “Any clinic visit past year”, so that the two appear not to be directly comparable. I understand fully that the results yield different results, with the pictorial diary suggesting that lower user-fees led to greater utilization and the household survey method showing no result—that’s fine. At the same time, if this can be shown on precisely the same measure, it would make the point that the measurement error is not separable from the treatment effect a much stronger point. I would have preferred a full-blown regression specification with child fixed-effects, where the randomization is interacted with the data collection method to
look at all effects (and the statistical significance) in a single regression framework. This can be done with and without additional adjustments.

We thank the reviewer for this suggestion. In Table 2 we continue to provide treatment effects for each of the different measures from the two methods of data collection. In Table 3 we then analyse the data in a single regression framework. We stack the utilization data from the two sources and run a regression in which we include a free care dummy, a household data collection dummy, and an interaction between the two. The coefficient on the interaction identifies the difference in the effect of free care between the two data sources. We include controls and cluster the standard errors at the household level. We carry out this analysis for utilization in the past month and for utilization with the extended reference period (i.e. six months for pictorial diary data and one year for household data). We recognize that the latter is not perfectly valid because the measures are defined differently between the two data sources – still though we think it is worth reporting.

We have run the regressions with child fixed-effects but it is not possible to include socio-demographic controls at the same time (due to perfect collinearity). Fixed-effects regression yield extremely similar results, as expected, but we have chosen not to report these results to keep things succinct (the journal does not appear to offer appendices). Instead we have inserted an endnote to this effect (page 10). We are happy to report the child fixed-effects estimates should the weight of opinion be in favour of doing so.

One further point, we have run all regressions with OLS to keep things simple and easily interpretable. The results remain very similar with logit / probit marginal effects but we thought we should bring this attention to the reviewers.

10. Discussion and other literature: 1. One of the reasons for the difference could be that the treatment effects and the bias in measurement are non-linear, which would break the required separability of the measurement error from the treatment effect. For instance, if under-reporting increases with higher morbidity at an increasing rate, then declines in morbidity alone could generate different coefficients in the treatment effect on reported morbidity and health seeking. This is what was found in Das and others (JDE), and the fact that the gradient of utilization with wealth was negative was one of the main findings there, leading the authors to suggest that the belief that the rich use health care more than the poor may have been directly generated by the way that data were collected.

This point is well taken. However, our understanding is that this is less of problem now that we use a standardized (binary) measure of utilization in Table 3. Is that correct? If not, we would be grateful if the reviewer could expand further on this point. Note that there is no difference in morbidity between intervention and control (from the pictorial diary data).


Thank you for this reference – which we now cite (page 5).

12. In the introduction, do the authors want to suggest that there has been recent work on recall, but less on different ways of data collection, noting that their method does not allow recall effects to be separated from data collection methods?
We agree, we are unable to separate recall from the method of data collection and have made this point as a limitation in the discussion (page 14). We highlight our contribution as one concerning data collection methods in the introduction (page 5).

If you have any further queries please do not hesitate to get in contact.

Yours sincerely,

Timothy Powell-Jackson and Evelyn Ansah