Reviewer’s report

Title: Antecedents of teenage pregnancy from a 14-year follow-up study using data linkage

Version: 1 Date: 22 September 2009

Reviewer: Julie Bayley

Reviewer’s report:

Review Summary
I feel this is a well constructed piece of research and adds considerable value to the body of literature on teenage pregnancy. The longitudinal data tracking and thus objective methodology is valuable and highlights some very interesting antecedents, many of which have not really been highlighted before. I offer some constructive comments below to improve the quality of the paper to support the value of the work. Yes, and I am particularly glad the authors spend a significant portion of the discussion on caveats and potential confounds / unmeasured variables. The results are extremely interesting, but as the authors say there are some findings which cannot be explained via this data alone (eg. link between parents smoking and teenage pregnancy). The authors explain these issues very well. The authors adequately cite the original survey, and give suitable context. They do not seem to utilise any other sources. The writing is very easy to follow, although a small amount of rephrasing would be very beneficial as detailed below.

Major Compulsory Revisions
None

Minor Essential Revisions
1. Please add a summary table of the sample demographics from the original survey and the new sample to allow the reader to understand the generalisability of the findings

2. I’m unsure of the ethics process – did you obtain consent to access records? Or is patient consent not needed to access records via the WADLU? Whilst the team were given an anonymised dataset, I would like to know whether the sample were aware of the way their data was being used, or if they were given the opportunity to consent, or if this was not required.

3. Given that the total sample size at 14 year ‘follow up’ is 155, please indicate whether this is a large enough sample to determine statistical significance

4. P8 line 1: ‘a de-identified confidentialised file’ – I’m not sure what the author’s mean by confidentialised? If it is anonymised (which I think is a preferable term for an international audience), then how is it also confidentialised? I would consider ‘confidential’ to relate to the process of use rather than the manipulation
of the data. What do the authors mean here by confidentialised?

5. P11 para3: “The proportion of girls who became teenage mothers...” - as a point of clarification, is this the proportion of girls who became pregnant and continued to parenthood rather than all those who got pregnant? If so, could the statement be rephrased to eg ““The proportion of girls who became pregnant and continued to parenthood...”

6. P12 para 2: I am slightly uneasy about the use of the word ‘if’ ie: “Some 22.3% of girls became pregnant as a teenager if their primary carer completed Year 9 or lower at high school”. Whilst the nature of the study allows more causative links to be posited, ‘if’ suggests the link is directly causative and implies that if the carer had completed post Year 9 education, they wouldn’t have got pregnant. Could the authors please rephrase to address this.

7. P12 para 4: This is a little arbitrary, but I don’t like the use of the phrase ‘the girl’ throughout this paragraph. This is because it changes the feel of the paper from large scale links between family context etc and teenage pregnancy to an individual case profile. Whilst I understand this is based on the hazards model, the phrasing doesn’t sit well among the remainder of the paper and from a non-academic viewpoint is almost accusatory. Could the authors rephrase by at the least taking out the repetition of ‘the girl’, and ideally finding a new way of phrasing this

Discretionary Revisions

8. I think overall the question is well defined, but I offer a slight caveat to this. There are two questions presented in the introduction:
   a. “We hypothesised that there are factors in early life that are predictors of teenage pregnancy”.
   b. “We investigated whether the observed associations changed with the age of the girl when the risk factors were assessed”

   Most readers in the topic area would accept (a) anyway based on the same issues the authors have mentioned in the introduction – eg. family context, parental relationships, social exclusion etc. I think (a) should be unpacked to show the unique selling point of the paper. The authors should rephrase it to give a sense of direction, or qualify what they mean by factors – do they mean factors which have not previously been highlighted? Or are they going to quantify the impact of known predictors? Or is the emphasis on early life? On reading the paper I feel they target all three of these, so I think this overly simple question undervalues the work. Question (b) is much clearer and I assume it is a secondary research question based on its position in the text.

9. Does the data allow the researchers to identify demographics of the sample at 14 year follow up? It would be very useful to determine if there has been any social mobility in the sample which might have skewed the results.

10. P18 para 1: “Girls are less likely to delay first sexual intercourse and use
contraception at first intercourse if they had not lived with both natural parents to sixteen years of age”. This is an akward phrase as it has a mix of positive and negative qualifiers and boolean terms. Given the necessity of the first statement, could the authors rephrase the second part to limit the awkwardness eg. change ‘if they had not lived’ to. ‘if they lived seperately’

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests