Author's response to reviews

Title: Explanations for female excess morbidity in adolescence: evidence from a school-based cohort in the West of Scotland

Authors:

Helen N Sweeting (helen@msoc.mrc.gla.ac.uk)
Patrick B West (patwest@msoc.mrc.gla.ac.uk)
Geoff J Der (geoff@msoc.mrc.gla.ac.uk)

Version: 2 Date: 31 July 2007

Author's response to reviews: see over
We would like to thank both reviewers for their very considered review of our paper, which we have now substantially amended. At the most general level, the emphasis of the paper is now ‘psychosomatic’, rather than ‘physical’ symptoms, and this is reflected in the change to the paper’s title. Following statistical advice from Geoff Der, who is now also included among the paper’s authors, we have changed the presentation of the analyses. In addition, we have increased both the Introduction and Discussion sections. We detail all the changes in our responses to the reviewers below. Please note that these are bulleted, following each of the reviewers’ points.

Torbjorn Torsheim Reviewer:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. Introduction. The manuscript would improve if a clearer statement of mechanisms was included. The authors consistently refer to associations, but without stating the mechanism. To me it is not clear why BMI should lead to more headaches in girls. There should be a clearer statement as to why self-esteem might explain gender differences in health. After all, the correlation between self-esteem and depressed mood tend to be very high, which could reflect conceptual confounding between these constructs.
   - We have expanded the Introductory section somewhat, and added further literature. This includes a reference to the close relationship between self-esteem and depression (page 6): “indeed it has been suggested that positive self-esteem and depression form opposite ends of a single continuum (Watson 2002)”, and an expansion of the paragraph in relation to BMI.
   - We acknowledge that there is no reason why each explanatory factor should relate to, or contribute to sex differences in every health measure and have added the following (within Hypothesis 2, page 9): “Thus, while we might, for example, expect physical activity to be associated with depressive mood, there is no a priori reason to think it would be related to headache. However, for completeness, we begin by examining the relationship that each explanatory factor has with each health measure, so contributing to the rather scarce (Vingilis et al, 1998) literature on predictors of adolescent self-rated physical health.”

2. Measurement. For some of the scales, notably the Masculinity scale, the operationalization is based on limited information. It is quite understandable that space limitations put constraints on the number of items included, but the authors should address whether this has any impact on the validity of the scale used to measure masculinity and femininity. In general, authors should include more about measurement validity and reliability, since these are important issues for making inferences from the data.
   - Within ‘Methods’, the section on the measurement of gender role orientation has been expanded somewhat and notes that the items included were “those with particularly high weightings on factors corresponding to femininity and masculinity on the adult Bem Sex Role Inventory (Bem 1974; Hunt et al 1996) and versions for children and young people (Boldizar 1991; Thomas et al, 1981)”.
   - We now include the ranges, means and SDs for all scales included in the analyses, and note that the Alphas for the masculinity and femininity scales are below the conventionally desirable level.
Finally, within our Discussion, we have extended our consideration of the results in respect of the sex by masculinity interaction on depressive mood (pages 18-19) and, within a new section on the study’s limitations, include a discussion of measurement error (page 19).

3. BMI comparison across gender is problematic, since any given BMI would reflect different body composition for boys and girls. An alternative would be to use standard age and sex adjusted classification of overweight and obesity.

- We recognise that total BMI represents both fat plus fat-free mass and that an age and sex adjusted obesity classification has a number of advantages. However, the consensus from studies which have examined the association between BMI and adiposity (e.g. Rodriguez et al, *International Journal of Obesity*, 2004, **28 Suppl 3**:S54-58; Wickramasinghe et al, *Annals of Human Biology* 2005, **32**(1):60-71; Zimmermann et al, *American Journal of Clinical Nutrition* 2004, **79**(5):838-843; Frontini et al, *Journal of Clinical Epidemiology* 2001, **54**(8):817-822) is that BMI is a reasonable measure of adiposity. There is also some evidence, which we now cite in our Introduction (page 7) that both underweight and overweight may be associated with well-being, but in different ways for male and female adolescents. We therefore felt it was important to capture the full spectrum of BMI, rather than simply to focus on obesity. However, in addition to reporting no association between BMI and any of our health measures, we now also note that obesity as defined according to UK90 cut-offs was similarly unrelated to any health measure (page 15).

4. Sampling. The sample is school-based, but it is not clear whether the sample was clustered. A clearer description of the sample is needed to consider whether a design effect might be present. The analysis does not seem to take into account potential design effects. Such design effects might be ignorable, but should be evaluated.

- We now include a paragraph describing our schools-based sample selection (page 10). Previous experience with the dataset had led us to believe that design effects would be negligible, as suggested by the reviewer. In the sample analysed here, intraclass correlations are 0.01 for sex, and range between 0.002 and 0.01 for the outcomes. It is therefore unlikely that the variance inflation would materially affect the results, especially with the more stringent significance level now used. We have added a note to this effect in the text (page 14).

5. Based on interaction analysis, the authors modelled gender differences in subcategories of potentially mediating influences. This resulted in separate analysis of gender differences in dizziness among non-smokers and smokers. While such a procedure could be justified in terms of achieving homogeneity in odds (statistical reasons), it is questionable whether this procedure makes sense conceptually. From apriori knowledge, and the introduction part, I would not expect a different pattern of gender differences among smokers and non-smokers. This procedure seems to be data-driven, with the danger of capitalising on chance. The authors need to justify their approach clearly.

- Following suggestions from the other reviewer, the interaction analyses, conducted in order to identify what we describe as sex differences in ‘susceptibility’ to each factor (corresponding with recommendations for the identification of effect modifiers) are now conducted last. This means that separate analyses of factors explaining sex differences in dizziness among smokers and non-smokers, and of factors explaining sex differences in depressive mood among those with ‘high’ and ‘low’ masculinity are not now included in Table 3. Further, the decision to adopt a more conservative .01 significance level, due to
the large number of separate analyses (as suggested by the other reviewer) mean that the
sex by smoking interaction on dizziness is now not an issue.

6. Table 3 includes a computation of "% female excess explained". This could be misleading,
since the logistic OR is a nonlinear transformation of the log odds. An OR drop from 3 to 2.5
and a further reduction from 2.5 to 2 is not on a linear scale.
• Although, as this reviewer notes, this measure would give different results on the log odds
scale, we do not believe it is seriously misleading. Results are expressed as odds ratios
throughout the paper and the measure is intended as an aid for the reader in identifying the
more important explanatory factors. Any reader who wished to make the comparison on
the log odds scale would be able to as the relevant odds ratios are all given.

7. The discussion is clearly written, but highly descriptive. Main results are mentioned and
emphasised, but not discussed at any length. A clearer contextualisation of the findings would
improve the manuscript. For example, on p.13 the authors note that among females,
masculinity was positively associated with depressive mood but among males it was
negatively associated. According to the authors this finding runs counter to other studies.
Some reflection about how this finding should be interpreted, and in what respect it seems
deviates form other studies should be included.
• The Discussion now focuses less on summarizing the results, but has been expanded to
include a fuller discussion of the results in respect of masculinity and depressive mood and
to cover the study’s limitations.

8. Much of the discussion deals with the lack of explanation of gender differences. Clearly
gender differences, as other differences in health, are multi determined. If we in addition
consider measurement error, there is every reason to expect that one will not be able to
explain gender differences completely. The impact of measurement error needs to be clearly
stated.
• We agree absolutely with the reviewer on this point, and within the Discussion’s section on
limitations (page 19), we now note that: “The majority of our ‘explanatory’ variables, and all
the health measures were self-report, and so likely to include measurement error to a
greater or lesser extent. Measurement error in a mediator variable tends to result in under-
estimates of its effect, which is ‘not a desired outcome, because successful mediators may
be overlooked’ (Baron et al, 1986, p.117). Given this, it is likely that female excess
morbidity could never be completely and satisfactorily explained by a study such as ours.”

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a
term, which the author can be trusted to correct)

The term ‘depression’ is used throughout the manuscript, but the measurement instrument
deals with depressive mood. Depression is a clinical term while depressive mood describes
variation in affective state.
• The reviewer correctly points out that the authors of this scale referred to it as measuring
‘depressive mood’ and, as he recommends, this is how we now also refer to this measure.
However, we continue to use ‘depression’ in the Introduction if / when this is the term used
by literature to which we refer.
Discretionary Revisions (which the author can choose to ignore)

The authors consistently write ‘Explain’ with a hyphen. Although it is good to express some awareness about the limitations related to explanation, the usage of the term can still be qualified in a simple sentence, rather than hyphenating it on every occasion. (e.g. We use the term "explanation" in a strictly statistical sense of the word.)

- We have now removed the quotation marks around every occurrence of ‘explains’/‘explanation’ and note in the Abstract and Hypotheses that explanation is statistical, with a further reminder to readers within Results (page 16), that “Table 3 shows the percentage of the female excess explained (in statistical terms)”.

John Cairney Reviewer:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

(1) I am not at all convinced that this paper explores both physical and mental health outcomes. Headaches, stomach/ache and dizziness could easily be reactions to stress, and could be combined with symptoms of depression to form an overall construct of psychological distress (a measure and concept found commonly in the literature). The authors need to make a much more compelling case for these symptoms as markers of physical health, including an analysis to see how distinct these symptoms actually are from depression (correlation, factor analysis etc). If these are not distinct (or overlap considerably), I recommend analyzing either a single construct – distress - or simply restricting the analysis to depression only.

- We agree with the reviewer that it may be unwise to describe headache, stomach problems and dizziness as ‘physical’ symptoms, and recognise that such symptoms are the ones often described as ‘psychosomatic’ and assumed to have a substantial psychological component (Pennebaker, 1982). We have therefore amended the title, and the paper itself, to reflect the fact that the symptoms on which we have chosen to focus are better described as ‘psychosomatic’. We include a discussion of this point within the Introduction (Page 4), note the association which each has with depressive mood in this sample (page 8), and suggest that “studies which seek to account for a female excess in ‘physical’ symptoms of a less clearly psychosomatic nature would require different sets of factors” (Discussion, page 21). However, it is interesting that other authors have defined symptoms that are possibly even more ‘psychological’, for example, sleep problems, tiredness, lethargy and sore muscles, as ‘physical’ (Eiser et al, 1995).

- We also acknowledge that an alternative to analysing individual symptoms would have been to create a symptom score, or a variable representing ‘high’ symptoms. Indeed, this is an approach that we seriously considered early on in our analyses for this paper. However, we have not adopted this method for two reasons (noted on page 9). Firstly, we wished to explore whether different factors were required to explain the female excess in each different health measure. Secondly, it is possible that factors associated with a general propensity to report high numbers of physical symptoms differ from those associated with reporting any one individual symptom.
The statistical analysis is problematic, largely because it departs quite significantly with common, well-established methods for estimating mediating and moderating effects (see Baron and Kinney, 1986; Mirowsky 1999). This is a significant limitation with this work. First, it is more conventional to begin with mediational analyses, not moderating effects as presented in Table 2. Perhaps because of the approach taken by the authors, this table is difficult to follow and requires greater detail to describe what is actually being presented. For example, do the Odds Ratios, for example, represent main effects, or do they represent the interactions? It is not clear. (On that, this is cross-sectional data. Avoid the use of the term relative odds, which implies Relative Risk, and is not appropriate given the data) Moreover, it seems as though the authors have simply run a series (at least 28) of simple logistic regressions, testing for gender and gender by risk factor interactions. Yet, there is no adjustment for multiple testing (i.e., accounting for the fact that by running so many tests, you significantly increase the probability of finding a significant effect, when one does not truly exist). The effect of gender by current smoking status, for example, would not likely survive a Bonferroni (or even a less conservative) adjustment. I think this problem, and much greater clarity of the results, could be achieved by taking a different approach to the data. First, I would run a series of staged regressions (as done in Table 3), for each outcome. This will provide a test of mediating effects. Please label the principal or focal variable “gender” so the reader can follow the analysis better. Next, I would force enter a set of interactions (between gender and each risk factor) into each of the four models (headaches, stomach, dizziness, depression – assuming you still have 4 outcomes, see point #1). Make sure all of the risk factors are in the model before interactions are included. You can do this two ways (force enter all interactions, or run separate models for each interaction – you will need to make an adjustment for conducting multiple tests, e.g., \( p > 0.01 / \# \text{ of tests} \)). I think once you have done this, you may only have one interaction – gender by masculinity – to worry about it. To interpret the effect, you will need to estimate the simple slopes for both boys and girls separately (see Aiken and West, 1991). Until this analysis is conducted in this way, it is difficult to evaluate the discussion/conclusion in its present form.

We would like to thank the reviewer for these suggestions. Within our Introduction, we now specifically link the section which refers to Rutter et al’s criteria for the identification of factors underlying sex differences in a disorder to the concepts of mediating and moderating variables, as discussed by Baron et al (1986) among others. We have also ordered our analyses so as to investigate the moderating effects of sex on each potential explanatory factor (the interaction analyses) last. This means that while Table 1 is unchanged, Tables 2 and 3 are simplified. Table 2 does not now include the significance of the interaction between sex and each potential explanatory factor on each health measure, and Table 3 does not include two columns for dizziness (smokers vs non-smokers) or depressive mood (‘low’ vs ‘high’ masculinity). Finally, the results of the interaction analyses are just described in the text. For these analyses, as advised by the reviewer, we entered the main effects of all the explanatory factors, plus the interaction between sex and each factor singly (described within the Analyses section, page 13).

Following the further suggestion, to adopt a more stringent significance level of .01, because of the number of analyses conducted, we were left, as predicted by the reviewer, with only one significant interaction, that between sex and masculinity on depressive mood. Since the remainder of the results are expressed in terms of odds ratios, we present the OR of a female excess in depressive mood for those of ‘low’ and ‘high’ masculinity (both unadjusted and after adjustment for the factors found to have been associated with depressive mood). These ORs demonstrated a female excess among those of ‘high’
masculinity only, evidence that higher masculinity was associated with depressive mood in females but not males (pages 16-17).

- We have amended the titles of Tables 2 and 3 from ‘relative odds’ to ‘odds ratios’.

(3) There are no psychometric or descriptive data (mean, sd’s range) provided for any of the scales that are used in this study. This must be included; especially for measures such as gender orientation, which appear to have been created by the authors.

- As noted in our response (2) to the other reviewer, we now include the ranges, means and SDs for all scales included in the analyses, and the Alphas for our masculinity and femininity measures.

-----------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

(1) Page 4, First paragraph. There are recent papers that support age 15 as a critical age for examining gender differences in depression which should be cited:
The last paper in particular shows gender differences in depression to emerge by age 14-15, using longitudinal data from 3 different countries, including the UK, and multiple measures of depression.

- Thank you for these suggestions; we have included the Wade et al reference in the very first sentence of the paper.

(2) Why use BMI as a continuous measure, when standard cut-offs for children and adolescents have been created to measure overweight and obesity? (see Cole et al., 2000).

- Please see our responses to reviewer 1, point (3).