Author's response to reviews

Title: Fertility in four regions spanning large contrasts in serum levels of widespread persistent organochlorines: a cross-sectional study

Authors:

Gunnar Toft (gultof@as.aaa.dk)
Anna Axmon (anna.axmon@med.lu.se)
Aleksander Giwercman (aleksander.giwercman@kir.mas.lu.se)
AneMarie Thulstrup (anemt@as.aaa.dk)
Anna Rignell-Hydbom (anna.rignell-hydbom@ymed.lu.se)
Henning Sloth Pedersen (hsp@natur.gl)
Jan Ludwicki (K.LUDWICKI@PZH.GOV.PL)
Valentina Zvyesday (dimusic@ic.kharkov.ua)
Alex Zhinchuk (dimusic@ic.kharkov.ua)
Marcello Spano (spanomrc@casaccia.enea.it)
GianCarlo Manicardi (manigi04@mail.unimo.it)
Eva Bonefeld-Jorgensen (ebj@mil.au.dk)
Lars Hagmar (lars.hagmar@ymed.lu.se)
Jens Peter Bonde (jpb@as.aaa.dk)

Version: 2 Date: 26 September 2005

Author's response to reviews: see over
Title: Fertility in four regions spanning large contrasts in serum levels of widespread persistent organochlorines

Authors: Gunnar Toft, Anna Axmon, Aleksander Giwercman, AneMarie Thulstrup, Anna Rignell-Hydbom, Henning Sloth Pedersen, Jan Ludwicki, Valentina Zvyesday, Alex Zhinchuk, Marcello Spano, GianCarlo Manicardi, Eva Bonefeld-Jorgensen, Lars Hagmar and Jens Peter Bonde

Author response to points rasied by the reviewers (comments in yellow and actual changes of the manuscript in blue)

Russ Hauser:
This is a well-written manuscript that explored the relationships between persistent organochlorine pollutants (POPs) and regional differences in several indicators of reproductive health in cohorts from four countries. The methods are generally well described. The authors present interesting data on the demographic characteristics of the four cohorts, the POP exposure levels, and data on measures of reproductive fitness (time-to-pregnancy/fecundity, semen quality). The authors reported between country differences in markers of reproductive health that they hypothesize may be related to between country differences in levels of POPs (PCBs and DDE). The authors offer alternative explanations for these associations, which include potential selection and confounding bias. For instance, there were large between country differences in potential confounders such as lifestyle factors, sexual behavior (e.g., frequency of intercourse for the TTP analysis) and medical histories (e.g., prevalence of urogenital diseases) that may confound the results. In addition, for the semen analyses there are likely differences in raters across countries, potentially dampening associations of POPs with semen quality, if a relationship did exist.

Thanks

Major Compulsory Revisions
1. Brief mention is made of eligibility/exclusion criteria for the subjects within each country. Further details should be provided to allow readers to determine if there are important differences in the selection criteria between countries. A brief comment is warranted on whether the recruitment setting (e.g., clinic versus hospital) may affect the characteristics of the study population within each country.

The following sentence added the methods section:

The recruitment settings are not expected to affect the characteristics of the four study populations. For instance, none of the clinics or hospitals differentially received women with high-risk pregnancies. However, the use of an earlier cohort for the Swedish sample resulted in higher age of the fishermen.

2. There are marked across country differences in the mean number of days (232 to 166 days) the women were pregnant at enrollment. The potential for these differences to introduce bias needs to be discussed. For instance, in countries where women were recruited later in pregnancy there may be more 'missingness' of women or couples with early pregnancy loss. If the early pregnancy loss were related to fecundity or semen quality this may introduce differences in these measures across countries.
Agree. We have added the following to the discussion:

Time to pregnancy studies of time trends and regional differences are probably exceptionally susceptible to bias that may be difficult to identify (Markku S, Juul Jensen). Although the recall of the time taken to conceive is expected to be highly valid in the pregnancy based studies, across region differences in the average weeks (23 to 33) the women were pregnant at enrollment could cause biased comparisons of TTP if pregnancies that were terminated by a spontaneous abortion were not included on equal terms in the four regions. Earlier studies indicate links between subfecundity and spontaneous abortion (Basso and Olsen 2000). Nevertheless, we believe this is a minor problem in this study because more than 86% of women were enrolled after the 12th week of pregnancy and thus is at low risk for abortion. However, if exposures in addition to subfertility first of all causes early abortions out study would not pick up such effects.

Minor: Please present values in weeks (traditional gestational length metric) and provide a measure of variance around the central tendency of the distribution, such as median and 25th and 75th percentiles if distribution is skewed.

OK, weeks rather than days, measures of variance also provided in the methods section

3. On page 17, borderline p-values (0.08 and 0.07) should not be used to dismiss potential differences between regions. There should be more discussion of these results, specifically the sperm morphology differences.

We have changed the text to emphasize magnitude of difference rather than p-values:

We did not observe obvious differences between regions of crude values (p=0.08), sperm counts adjusted for trivial factors as period of abstinence (p=0.22) or sperm counts adjusted for urogenital infections (p=0.13).

The adjusted geometric mean percentage of men with morphological normal sperm differed up to 15% among regions with lowest values in Warsaw and highest values in Sweden (Table 7).

4. Page 18, although there were regional differences in fecundability, the statement that it was related to population DDE levels and not PCB levels may be too strong. There are only four data points when country is used as the indicator of exposure.

We point to the fact that at the regional level the FR distribution follow the DDE distribution (4 data points) but not the PCB distribution. Whether associations are causal is extensively discussed in the following paragraphs.

5. The statement that men in Warsaw had low levels of CB-153 and DDE but could have higher exposures to other compounds should be expanded. Is there evidence of this and if so to what exposures? Is this based on data or observations not presented here?
Good point. Unfortunately we have no data to substantiate this statement. We have at this stage no other measurements of any compounds besides cb-153 and DDE but hope to get funding to perform analyses on subsets of people from the four regions to examine regional exposure profiles. This limitation is now clearly acknowledged:

However, at present we have no data to indicate that people in Warsaw are more or less exposed to the wide range of xenobiotics that potentially could interfere with fertility.

6. Page 19-20, the discussion of methodological differences affecting study results is well-done but needs to be further emphasized since it likely accounts for some of the across country differences in markers of reproductive function. Can some of this discussion be moved to earlier in the discussion section and specifically mentioned in the abstract?

We are reluctant to move the methodological limitations higher up in the discussion since that will completely change the logic – first we provide the main findings, then we discuss these findings in light of the hypothesis (that xenobiotics are to blame) and then we discuss other interpretations. The conclusion of the abstract has been changed:

**Conclusion:** We observed regional differences in time to pregnancy and sperm motility that may be related to regional differences in POP blood levels but other interpretations are also plausible. Across region differences in access to safe contraception and in prevalence of contraceptive failures is likely to bias comparisons of time to pregnancy, in particular.

7. The tables are generally well-designed and present the data well. The following are brief comments that require minor discussion or minor revisions.

8. Table 1 shows marked differences in participation rates and contraceptive failures across countries (up to 50% in Kharkiv). This raises the concern with bias (i.e., selection and confounding) in the analyses and makes the between country results difficult to interpret. Was participation rate considered as a predictor in the models? This may allow further exploration of whether it was predictive of the outcome of interest (TTP or semen analysis results).

We fully share the concern with respect to biased comparisons of time to pregnancy and believe that this aspect in response to the above remarks has been given even stronger attention in abstract and discussion. Unfortunately, it is not possible to adjust for participation rate as this is a region characteristic and not an individual characteristic. We do not have participation rates at any level lower than for the entire region (except in Greenland where the participation was high in all districts).

9. In Table 1, the percent participation in the semen sample collection row should be added to the row below number of semen samples collected.

Done.
10. Tables 2 and 3 showed marked across country differences in demographics, especially for percent of current smokers (note: there are also likely large differences in number of cigarettes smoked among current smokers), urogenital infections and percent daily intercourse. These marked across country differences again raises issues of concern with inability to adequately control for potential strong confounding by these factors in the between country analysis. This should be commented upon in the discussion.

The following has been added to the end of the discussion: Although all studies were designed and executed according to an agreed uniform research protocol, the marked across region differences in demographic factors as age, lifestyle factors as smoking, urogenital infections, contraceptive methods, reproductive behaviour, periods of sexual abstinence and season of sample collection call for an adequate control for potential confounding by these factors in the between regions analyses. We chose the change-in-estimate method to keep factors in the models regardless of p-values (Greenland) and known strong determinants as age and period of abstinence were compulsory in all models. Even some residual confounding cannot be ruled out, the main threat to the internal validity of this study is probably strong selective forces that cannot be adjusted for in the analyses.

11. Table 4: The large differences in contraceptive failure rates across countries raise concern with bias introduced by excluding unplanned pregnancies due to across country differences. This should be further emphasized in the discussion. Cf above, moreover emphasized in the revised abstract. Se also discussion p 20-21 and 24-25

12. Table 5: Minor point: percentages are row percentage totals (thanks, changed). FR changed after adjusting for current female smoking and daily sexual intercourse. This raises the concern that finer adjustment (number of cigarettes smoked or a continuous measure of sexual frequency) may further alter FR. Please comment.

Current smoking among the women reduce fecundability with some 20-30% in all regions – which is as expected. Therefore adjustment based upon crude categories (smoking yes/no) seems to work well. We did also models based upon number of cigarettes/day and obtained risk estimates in the same range.

13. Table 7: marked differences are noted in potentially important confounders, e.g., abstinence time, season when sample collected, infections, medication use. This should be commented upon and discussed in the text.

These factors are discussed extensively, please consider page 22-23

14. Table 8: I was unable to view on screen or print out. The table is corrupted in my version.

Hopefully the next version works.

15. Figure 11: FR figure, consider the use of log scale for FR.
We do not understand this suggestion? But the figure has been omitted as suggested by the second reviewer?

Minor Essential Revisions
1. Page 6: define short delay between enrollment and semen collection OK, done

2. Table 7: Spelling: Fever not fewer, thanks, corrected

Discretionary Revisions
Page 23- Future methods that will be used for directly testing the associations between POPs and markers of reproductive fitness should be briefly mentioned.

OK, comes here:
In addition to TTP and semen quality these studies will also include indicators of sperm chromatin structure, DNA damage and apoptosis, sexual hormones, X/Y chromosome ratio in sperm and epididymal and accessory gland function.

2. Table 7: To provide insight into variance about the central tendency of the distribution, present 25th and 75th percentiles rather than 5th and 95th.

We provide mean (SD) as well as medians and believe it is of interest to know about the range of values and would like to stic to the 5\textsuperscript{th} and 95\textsuperscript{th} percentiles.

Tina K. Jensen:
General comments
This is a very interesting manuscript dealing with problems of general public interest. The authors have collected fertility data and semen quality among couples from four polluted areas. This must have been a demanding task and the results are interesting. The authors try to relate geographical and inter-country differences in POP exposures to fertility and semen quality. This is building on an assumption of a relationship between TTP, semen quality and POP exposure even though no data showing any relation between POP and TTP and semen quality are presented. If no association is found other factors may explain the observed differences in TTP. The manuscript is long and the TTP data are hardly comparable. Differences in unplanned pregnancies, participation rate and contraceptive practices make it difficult to compare TTP across countries. Especially the Swedish data differs and they are collected with another purpose dealing with TTP retrospectively collected data from fishermen’s wives. I am therefore not sure that the Swedish data are comparable to the others and suggest that they are excluded even though the authors adequately discuss these biases.

Swedish data: we share the concern although we emphasize that the Swedish data were collected with the same purpose and with minimal adaptions used the same research tool. We are reluctant to omit these data because they were protocolled and should be reported for completeness. Moreover, we stress strongly the caution that is needed in interpreting of the data – not only the Swedish but all regional comparisons of TTP. From that point of view we think these data add to the limited database on TTP in various settings.
Introduction
The introduction is long and it is not necessary to start the introduction with the DBCP story as these were occupationally exposed. Likewise, the exposure assessment difficulties in epidemiology do not need to be addressed in the introduction. In the second paragraph the authors conclude that no association between maternal estrogen levels and congenital malformations in male reproductive tract and semen quality. The reference (5) is however only to POP not estrogens in general.

We have omitted the paragraphs on the historical context and the introductory remarks on exposure assessment and TTP methodology – as well as other information. Hereby the introduction has been reduced by by 50%. One reference was missing (Storgaard et al 2005). Has been inserted, thanks.

Materials and methods
This is also a long and very detailed section but adequately describes differences in data collection. It may be shortened for example the section on semen quality may refer to a previous paper.

As suggested we have reduced the section describing the semen analysis with a reference to a previous paper. Otherwise we acknowledge that the methods section is rather detailed – reflecting that we are presenting TTP as well as semen data from 4 regions in one paper and notice that the first reviewer ask for additional details.

Results
As mentioned in the general comments, I suggest excluding the Swedish data. TTP as well as semen data? Please see our response above.

There are 8 tables which is too many. I suggest combining table 1, 2 and 3 since they deal with differences between countries.

We have combined table 2 and 3 into one Table 2 but are reluctant also to include Table 1 which essentially is a flow chart describing the recruitment that results in the population which is characterized in Table 2 (and former 3). Moreover, we have omitted Figure 1.

In table 2 female smoking and alcohol consumption is shown, is that during pregnancy or TTP?

In the method section the following is included:
Information about time varying exposures as tobacco consumption and intake of alcoholic beverages was given with reference to the date when the couple started trying to become pregnant.

The differences in parity between countries makes it almost impossible to adjust for, it would therefore be interesting to see the TTP results for primiparous women across countries especially since POPs are strongly related to parity.

We have performed analyses only including primi-parae:
Similar findings were obtained when TTP values were censored after 18 and 24 months rather than after 12 months and in analyses only including the first pregnancies except that the latter analysis revealed a reduced fecundability among fishermen’s families [age and smoking adjusted FRR 0.58 (95% CI 0.35 – 0.97)]. However, this finding is based upon only 29 last planned pregnancies among Swedish women.

Likewise for women using similar contraception, since there are large differences in contraceptive use and diseases in diseases in reproductive organs across countries (Table 2 and 4).

Agree, additional analyses have been performed and the results have been added to the results section:

Methods:
We also performed sensitivity analyses that (i) used censoring after 18 and 24 months, (ii) included couples that became pregnant in spite of use of contraception (TTP assigned 0), (iii) included first parity pregnancies only, (iv) and that included couples than discontinued non-oral contraception only.

Results:
Similar findings were obtained when TTP values were censored after 18 and 24 months rather than after 12 months, in analyses only including couples not using oral contraception and in analyses only including the first pregnancies except that the latter analysis revealed a reduced fecundability among fishermen’s families [age and smoking adjusted FR 0.58 (95% CI 0.35 – 0.97)]. However, this finding was based upon only 29 pregnancies among Swedish women. Inclusion of couples that became pregnant in spite of contraception by assigning these couples a TTP value of 0 resulted in weaker differences between regions.

Why were FR differences not adjusted for differences in contraceptive practices (table 5)?
They were, explanation given in the footnote to Table 5:
Female urogenital disease or infection, low and high body mass index and use of oral contraceptives did not change at least one of the risk estimates with at least 10%.

Table 4 deals with a central problem in the TTP data. Sensitivity analysis including accidental pregnancies may be interesting to perform.

Agree, additional analyses have been performed and the results have been added to the results section, as above.

Table 6 may be excluded and the text may just mention that FR was lower among couples providing semen quality.

Disagree. Selection bias is a prominent and acknowledged problem in cross-sectional semen studies that seldom can be addressed directly. This paper can and findings are in our opinion important to give a rather definitive answer to this important question. Therefore we hope that the table can indeed be included.
Figure 1 a shows male serum DDE levels in the four countries and could just be added to a table. What were the levels of PCBs across countries? Figure 1 b shows FR across countries and is seen in table 5 and may therefore be omitted.

We omit the Figure although we think it illustrates a striking observation that may be difficult to comprehend that clear form the text. DDE and CB-153 levels are mentioned in the text.

Discussion
The discussion adequately deals with the possible biases mentioned above but it is long and after reading it, the reader is confused and left with the question off the added value of combining these data.

We have amended the discussion, please se comments to the first reviewer. And we have made considerable reductions of the original text so hopefully the discussion is easier to comprehend.