

Author's response to reviews

Title: A cohort study of in utero polychlorinated biphenyl (PCB) exposures in relation to secondary sex ratio

Authors:

Irva Hertz-Picciotto (ihp@ucdavis.edu)
Todd A Jusko (jusko@u.washington.edu)
M Judith Charles (jusko@u.washington.edu)
Rebecca J Baker (jusko@u.washington.edu)
Jean A Keller (j.keller@murdoch.edu.au)
Teri A Greenfield (stgreen@omsoft.com)
Eric J Willman (eric.willman@ecolab.com)
Stuart W Teplin (steplin847@nc.rr.com)

Version: 2 Date: 7 May 2008

Author's response to reviews: see over



DEPARTMENT PUBLIC HEALTH SCIENCES
DIVISIONS OF ENVIRONMENTAL AND OCCUPATIONAL HEALTH AND EPIDEMIOLOGY
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-8638
VOICE: (530) 752-3025
FAX: (530) 752-3239
Email: ihp@ucdavis.edu
<http://www-epm.ucdavis.edu/>

May 7, 2008

Dr. David Ozonoff, MD, MPH
Environmental Health, Editor
BioMed Central Ltd
Middlesex House
34-42 Cleveland Street
London W1T 4LB, UK

Dear Dr. Ozonoff:

Thank you for the opportunity to revise our manuscript "A cohort study of in utero polychlorinated biphenyl (PCB) exposures in relation to secondary sex ratio" (MS: 1431782571727880). We appreciate the opportunity to respond to the insightful comments from the reviewers.

Based on the reviewers' comments, it is clear that we did not adequately explain our weighting procedure. Accordingly, we have updated our manuscript with regard to the sample weighting/bias issue, and the other concerns raised by the reviewers. We believe the manuscript has been much improved by this process. Please find attached the point-by-point responses to issues raised by the reviewers.

We look forward to further correspondence from the Journal.

Sincerely,

A handwritten signature in black ink, appearing to read "Irva Hertz-Picciotto".

Irva Hertz-Picciotto, PhD

Professor
Department of Public Health Sciences
Deputy Director, Center for Children's Environmental Health
Chief, Division of Environmental and Occupational Health
University of California
TB #168
Davis, CA 95616

Attachment: responses to reviewers

Referee #4:

The problem is that the selection criteria may well have introduced selection bias. Table 1 includes a long list of reasons for exclusion from the 5-year cohort. (Mother is an unmarried minor, no interview completed, blood was drawn prior to second trimester or after 3rd trimester, not a singleton birth, Raven or Peabody cognitive test score missing, hearing tests were not done in both ears, confounding anomalies present (what are these?), mother is deaf, congenital rubella, mother took a drug containing iodine, thyroid drugs were taken within 60 d of blood draw, gestational age is less than 244 or greater than 351 days (or unknown), no current address is available, does not reside in 1 of 8 included counties, only one sibling in family can be included; other siblings dropped.

The largest exclusions were made for: 1) unmarried minors or if no interview was completed (17%); 2) subjects who did not have a current address (14%); and 3) subjects who did not reside in a subset of counties (10%). To have introduced selection bias, the above factors would have to be jointly associated with PCB exposure and infant sex. It is difficult to see how these selection factors would be jointly associated with PCB concentration and infant sex. In fact, very few factors have ever been identified that have much role in infant sex, but to be jointly associated with both infant sex and PCB exposure is even less likely. The remaining factors account for small percentages of the cohort and therefore any bias they could have caused is likely to be small.

Moreover, the child has to 1) have participated in the follow-up study at five years of age (meaning that they had to survive to five years and been able to participate in the follow-up exam).

The proportion of infants who survive birth only to die in the first 5 years of life is less than 2% (and was at the time of this study). We have added mention of the possibility of childhood mortality associated with PCBs, although given the small percentage of deaths and the low levels of PCBs, it is unlikely that an effect on children but not on fetuses would occur.

But then criteria for inclusion in the sample of 399 is more restrictive: Except for 68 children (the 17% random sample) the child had to 2) have gotten a low score on at least one of two cognitive tests performed or failed a hearing test. None of these would be exclusion criteria for a study conducted solely to examine (birth) sex ratio and PCBs exposure.

We note, to begin, that the 17% random sample was actually 159 children, not the value of 68 suggested by the reviewer (we sampled 17% of the eligible children, i.e., from the non-cases among the 1291).

Regarding the stratified sampling, this was accounted for by the analysis. Specifically, we applied weights to our regression analysis that reweighted our results as if children had been selected at random from the larger sample of 1291. This is the exact same approach used in virtually all large-scale survey research. Because this weighting was done, any bias introduced by our stratified selection was undone by the weighting, just as occurs when a stratified sample is obtained in a survey such as the NHANES and weights are applied to yield results for the full U.S. population. As a result, no selection bias occurs from the stratified sampling, as long as the analyses are weighted. This is similar to the issue of a matched case-control study: matching introduces a selection bias, but as long as the proper 'matched analysis' is conducted, there is no bias (See Rothman & Greenland for a detailed explanation of this point).

I am concerned because of the potential for selection bias introduced when selecting on factors that might, themselves, be related to PCB exposure (birthweight, gestational age, survival to five years, cognitive performance, and particularly when these effects may be sexually dimorphic. For example, if exposure differentially increased perinatal and infant death so that more exposed boys died between birth and the five-year follow-up (not implausible, particularly given these authors earlier paper reporting on a greater adverse effect of PCBs on male fetal growth), then higher PCB exposure would be associated with a deficit of boys (as was seen in this study. In fact, the male to female ratio in this sample (47:53) was quite a bit lower than that in the total cohort (51:49). Similarly, excluding premature infants (<35 weeks) may have had a greater effect on PCB exposed males. The requirement that children selected for inclusion have (for 83%) poor results on cognitive and auditory tests might also be related to exposure; Vreugdenhil (2002) found sexually dimorphic effects of PCB on children's play behavior. Widholm (2001) reported sexual dimorphisms in effects of PCBs on spatial cognition in rats.

We are in agreement with Dr. Swan that there is potential for selection bias if these factors are jointly associated with PCB exposure and infant sex. We agree that some results have shown sexual dimorphism, which is why we became interested in the question of sex ratio. So let's examine these factors to determine the likelihood of selection bias. We excluded very short and very long gestational ages, but our previous work in this same study clearly showed that whereas PCBs were related to intrauterine growth (adjusted for gestational age), they did not show any hint of an association with gestational age (Hertz-Picciotto et al 2005). Other studies have had similar results for preterm delivery. Note that we did not make any exclusion on birth weight. The proportion of infants that did not survive to 5 years, as mentioned, is likely quite small; generally speaking, boys die at a higher rate than girls. On the other hand, there is no evidence in the literature that PCBs at these levels were lethal. With regard to cognitive development, this was a stratification variable, not an exclusion criterion. As described above, the application of weights to the stratified sample of the 399 removes any bias introduced by the stratified sampling, and hence any associations with cognitive development will be removed in the weighted analysis.

The authors state "Normalized weights were used in order to represent the sex ratio distribution among all eligible children in the CHDS survey." In fact, using weights to adjust for sex ratio, when the outcome is sex ratio seems very questionable. How these weights were defined, and how they would eliminate selection bias (if they would) is completely unclear to me. I would also ask the authors to show the weighted and unweighted results.

Our wording was incorrect in the previous version of the manuscript. We have therefore corrected the wording and included additional explanation of the weighting procedure:

"In addition, normalized weights were applied to adjust for our stratified sampling procedure. By applying these normalized weights, which were inversely proportional to the sampling fraction in each stratum, to the regression analysis, the results are as if the sample of 399 children was drawn completely at random from the subset of 1291 children. Thus, any potential bias created by the stratified sampling procedure is removed. As in standard survey research, this procedure ensures that all parameter estimates are valid with respect to the larger sample or population from which the stratified sample has been selected, and eliminates any bias introduced by use of a stratified sample."

We used weights to represent the population from which we sampled, namely the 1291. The sex ratio in this group was slightly skewed towards more females than the original CHDS cohort; however, the comparisons we made were internal to the 1291, comparing sex ratio at different levels of PCBs.

We applied normalized weights which were based on the sampling fractions of children from the two strata: the 17% random sample, and the children selected who failed a hearing test/scored low on a cognitive test. This weighting procedure was done because our sample of 399 children was not random with respect to the 1291 children who met all eligibility criteria; rather, it was a stratified random sample from the 1291. To adjust for this stratified sampling, we reweighted our results to obtain the appropriate sex ratio as if the sample of 399 was drawn at random from the 1291. As in survey sampling, applying weights that are inversely proportional to the sampling fraction eliminates bias. Thus, showing unweighted results would be completely incorrect; again, as explained

above, it would be tantamount to showing a deliberately biased result from conducting an unmatched analysis of matched data, or to expecting the NCHS to not apply weights and instead report crude results from the biased samples obtained by stratification.

Interestingly, in commenting on Rogan (1999) the authors state: "The Yucheng group consisted of a select set of mothers who had registered with the health department as having been affected, and who had at least one live child in 1985. Not all affected subjects were registered, and some of the most severely affected children may not have survived: since either of these factors could have been related to the sex of the child, the sex ratio of the survivors could be highly confounded." This is precisely my point.

The Yucheng group suffered frank poisoning. It was *very* different from the CHDS cohort in terms of both the levels of PCBs and child survival: there is no evidence that PCB concentrations at the levels of the general population in 1964 caused any postnatal mortality in this cohort or anywhere else in the world. These concentrations were at least an order of magnitude lower than the Yucheng poisoned children. Additionally, persons in the CHDS cohort did not have to register with the health department as victims and suffer any associated stigma (in some Asian cultures, there is considerable stigma associated with developmental disorders or being a victim of this type of tragedy). If any child in CHDS died because of PCBs, their exposure levels would have been far outside the distribution examined in this study; the analysis we conducted provided information on PCBs in the observed range, yielding the appropriate relationship between prenatal PCB exposures and secondary sex ratio.

The heading for the sample used in this analysis in Table 2 is headed "Final random sample." While the 17% may be random, the entire sample of 399 is not.

It is actually a "stratified random sample." We have made the change.

Others studies finding associations between parental PCB exposure (predominantly paternal) and sex ratio suggest stronger associations when first exposure occurred before age 20. Cohn et al. found a similar result for DDT and breast cancer in the same Kaiser Permanente cohort used for this analysis. Cohen's analysis was possible because of the relatively clear onset of DDT use in 1945. While this was not the case for PCBs, their use undoubtedly increased greatly post-WWII. It would be interesting to see, to the extent possible, whether the association reported here is modified by maternal age.

The findings by Mocarelli et al showing a stronger effect on sex ratio from paternal TCDD exposures that occurred before age 20 are interesting and have been discussed in the manuscript. We tried fitting a model with an interaction by (categorical) maternal age, but we had trouble with model convergence (not uncommon in the log-binomial model). However, an interaction between PCBs and maternal age as a continuous variable was not statistically significant ($p=.25$), although power in this study for assessing interactions was small.

Referee #1:

The title, however, does not convey the main finding, but created possibly a curiosity so the reader is tempted to proceed with the manuscript. However, is secondary sex ratio studied when just live births are included?

Secondary sex ratio refers to the sex ratio at birth, as opposed to primary sex ratio which refers to the sex ratio at conception. We were unable to locate a reference, but a quick look at half a dozen published papers in this field indicates that authors are not including fetal deaths, since one of the main discussion points is whether sex ratio changes as a result of increases in fetal deaths.

In this study the use of a sample of the children with low score at cognitive tests at the age of five (+ 17 % random sample of all other) seem queer and might introduce bias. A random sample of the total cohort with all births (live and dead) would be better.

Because we reweighted our sample to the original proportions of children in the subsample of 1291, bias was not introduced by the oversampling of children with low cognitive tests. Our weights were calculated to be inversely proportional to the sampling probabilities, and therefore the biased result would have come from an unweighted analysis. (See responses to referee #4 above.)

Could PCB level in utero be associated with the low cognitive scores and hearing impairment at five years? If so, is the impact the same in male and females? More information ought to be presented for the study sample, especially the 17% random sample. How was the PCB level in the random sample compared to the rest?

This issue was addressed above, response to referee #4.

I have done some crude calculation based on the available information in the manuscript: Of the 399 children, 216 had low cognitive scores or hearing impairment, and 17% (183) were drawn randomly from the rest of the cases meeting study criteria (1291-216). For the sake of simplicity I say that the random sample had the same male proportion, as the cases meeting study criteria: 49% (90) male and 51% (93) female (should be higher male proportion since children with low scores seem to have a lower male proportion). If the ten percents with lowest PCB exposure belong to the random sample, the numbers would be the following, 2 by 2 table:

90 pct 10pct

High PCB low PCB

Crude odds ratio 0,86 (exact CI 0,40-1,75)

male 166 20

female 193 20

total 359 40

Or are just the 10% of the highest and 10% of the lowest PCB values included in the analyses in table 3? Unlikely since n is stated as 399. It is a need for a more detailed method description of comparison groups to avoid misunderstanding. Below calculation for the sample of the 10% of the highest and lowest PCB values. Below 2 by 2 table

10 pct 10pct

highest lowest PCB Crude odds ratio 0,82 (exact CI 0,31-2,15)

*male 40*0.46 40*0.49*

*female 40*0.54 40*0.51*

total 40 40

The referee has misunderstood the analysis. We used all participants for the analysis. We did not dichotomize or use subjects below or above a cut-off but rather fitted a regression model to the entire range of PCBs represented in the study. We then used the coefficients from the best fitting model to calculate the predicted sex ratio at the 10th and at the 90th percentile. Thus, to improve interpretability of the regression results, which provide the RRs in units per ng/ml of PCBs, we present the RR for a change from the 10th to 90th percentile of PCBs. In our analyses, we estimated the relative risk for each unit change in PCB concentration. The crude calculations by the referee reflect a misunderstanding of the nature of these data and the type of analysis that we conducted (i.e. a regression analysis on all subjects), which adjusted for the sampling as well as some covariates.

Might the PCB have a different cognitive effect on females than males? In the analyses it might be of interest to make an additional adjustment for the sub sample (low score or failed hearing screening and the random sample).

Yes, it very well may, which is why we have done exactly what the referee suggests. That is, we made an adjustment for the sampling; the proper adjustment for this stratified sampling is a weighted analysis.

In Reference 17 the authors write “Because we over sampled cognitive- and hearing-impaired children for this study, the current sample was not drawn purely at random. Therefore, weighted analyses that applied appropriate sampling weights and corrected for design effect were Performed.” Could something like this be done also in this setting?

This is exactly what was done in this setting.

PCB # 101 and # 156 have not been presented since more than 30% were below the limit of quantitation (LOQ). This might be sensible. However, it would still be interesting to know the tendency - was the male proportion reduced for these PCBs as well?

Approximately 61% and 46% of PCB 101 and 156 concentrations, respectively, were below the LOQ. At this level of “missingness”, significant bias results in the estimation of population parameters. Thus, any association, even if in the same direction as the other congeners, is likely to be highly suspicious.

Two articles from Europe from 2005 and 2006 by Tiido et al are not included in the list of references. They deal with human sperm Y:X chromosome ratio (1,2). One of the papers showed increased number of Y sperm in a group of highly PCB exposed men. It would be good if also these articles were included in the discussion, even if they are not treating secondary sex ratio.

We now include reference to this literature on sperm chromosome ratio to our manuscript.

The PCB analyses were done in batches (10-20), consequently 20-40 batches. Concentrations were adjusted for percent recovery for each sample- how was this done?

This is a standard procedure in analytic chemistry. During the extraction phase, some of the sample can be lost through evaporation, or other means. Within each batch of 10-20 samples, a sample standard is spiked with a known concentration of each PCB congener (including PCB 204). Then, to each serum sample, a known amount of PCB 204 is added. Once the samples are analyzed by the gas chromatograph, each sample can be adjusted to its original concentration based on the percentage of PCB 204 recovered in that sample, relative to concentration of PCB 204 recovered in the standard.

Specimen characteristics were included in the analyses- some more details on which characteristics are appreciated.

These are already described in the manuscript in the methods section under “Outcome and Other Variables:”

“Other potential adjustment variables for the multivariate analysis were specimen characteristics. These were: batch-date of laboratory analysis and an indicator for sample storage history (some specimens had been immediately shipped for storage at an NIH facility while others had been stored locally and later shipped).”

In the manuscript is stated that the PCB is stable, however, might the actual PCB serum level at blood draw have been lower than when it was analyzed due to for instance vaporization - after 40 year of storage at minus 20o Celsius? However, this might be the same for all samples and thus not introduce bias.

It's possible that some sublimation of sample occurred due to the long storage period, but it seems unlikely to be substantial. As mentioned in the manuscript, before analyzing any samples from our subjects, noncritical samples from the same cohort (i.e. of women in CHDS who were lost to follow-up) were analyzed to assess the integrity of the stored serum specimens. Organochlorine determinations demonstrated comparability with other historical samples, and lipids were in reference ranges.

In the abstract and in table 3 you write that sex ratios are compared between the group belonging to the 10th and 90th percentile of PCB. This also needs to be added under data analysis, and ought to be specified to avoid misunderstandings (as above).

We have added an explanation of this data analysis approach.

PCB concentration of the total maternal PCBs (page 12). Omit mean (confusing), and reorder the other: 10th percentile, median, 90th percentile.

We believe that the mean should be reported, and have retained the sentence as written.

Table 2. Cases followed up at 5 years (n=3412) Sex of infant F:50, M:50 after rounding.

We have made the change.

Table 2, The heading: “Final random sample” ought to be changed as some of these children are a random sample, the rest are not.

Same comment as above referee: We have corrected it to “Final stratified random sample.”

Has the manuscript dealt with secondary sex ratio when just live births are included? This ought at least be more thoroughly discussed.

This manuscript is based on a cohort of live births. The possibility of differential deaths by both sex and PCBs between birth and five years is now discussed (p. 17).

Referee #2:

The selection process involved selecting all those who scored low on at least one of two cognitive tests or who failed a hearing test, and only a 17% sample of those who did not meet these criteria. The authors indicate that this resulted in selecting more girls because of the relative performance on these tests of boys and girls. Because of this, if there is some association in their data between maternal PCBs and the cognitive and/or hearing tests, this would introduce bias. To deal with this, I think the authors need to do a weighted analysis such that the test scores (or at least their categorization of them) in their sample reflect those in the population from which selection into their study was made. The authors do do a weighted analysis, but this was with respect to weighting to achieve a sex distribution similar to the original cohort, which I do not believe would accomplish the same thing as weighting based on the cognitive and hearing testing categorization.

Our original submission was unclear. We did in fact perform a weighted analysis based on the stratified sampling design based on cognitive and hearing test scores, such that subjects from the different strata were reweighted to reflect the population of 1291 children. By doing this, we can interpret our results as if we had sampled the 399 children randomly from the subset of 1291 children. We have updated the methods section of our manuscript to better reflect what was done.

As an additional sensitivity analysis, the authors could take a random sample of 17% of the low scorers and look at the result excluding those low scorers not sampled this way. The power would be much less, of course, but the point estimate could be informative.

Because of the weighted analysis, it's as if we had sampled the entire group of 399 children randomly from the 1291. Thus, such a sensitivity analysis is not necessary and might not be meaningful given that it would be a smaller sample than what we actually had.

The cutoffs that define low for the cognitive tests (related to the sample selection) should be defined.

We agree and have added this information to our manuscript.

It would be nice to have a table showing the distribution of some of the covariates by PCB levels. In fact, showing how the cognitive scoring (and in particular whether the categorizations they used related to the sample selection) varies by PCB level would be very informative as this association would be important to possible bias from the sample selection procedure.

Since any bias from the sample selection procedure (oversampling low cognitive scorers and taking a random sample of the remainder) would be removed by applying weights, we don't see this table as providing much additional information.

The authors used an imputation process for PCB congeners below the level of detection. It has been shown that using the actual values obtained (even if below the LOD) leads to less bias. If these are actually available, rather than simply not obtained from the lab, it would be better to use them.

We used the best available data.

In the 3rd paragraph of the results, table 3 should be referenced.

We agree and have updated our manuscript.