An analysis of the reliability of self reported work histories from a cohort of workers exposed to polychlorinated biphenyls

Carl R Rosenberg

Abstract
An investigation was conducted to examine the reliability (reproducibility) of self reported occupational histories obtained from a cohort of 326 capacitor manufacturing workers who had participated in an epidemiological study relating health abnormalities to exposure to polychlorinated biphenyls (PCBs). For a subsample of the cohort (n = 164) in which occupational histories were obtained twice (in 1976 and 1979), reliability of cumulative exposure to PCBs ranged from 93.6% for the early PCB period (1947–70) to 95.7% for the late PCB period (1971–6). These respective reliabilities were lower, however, for workers who changed jobs often. Workers above the median value of a weighted job change index had early and late reliabilities of 89.9% and 83.6% respectively. Reliability is a relevant factor when calculating power or sample size during the planning stage of epidemiological studies, for interpretation or adjustment of estimates in the analysis stage, or for determination of study feasibility.

(British Journal of Industrial Medicine 1993;50:822–826)

For retrospective epidemiological studies in occupational settings it is often necessary to rely on self reported data to determine the exposure state of the study subjects. At issue in such cases is the validity (accuracy) and reliability (reproducibility) of the data. Of the two, the first is the more critical factor. Epidemiological studies that lack an accurate source of exposure data may be seriously flawed. The validity of self reported work histories was considered earlier in the cohort studied here and by other researchers for other populations.6–7

A high level of validity, although necessary for a good study, may not be sufficient to achieve that aim. Cumulative exposure estimates based on self reported work histories may closely approach the true values, but only if the histories are obtained at two or more points in time from each subject and the estimates averaged. This implies that an estimate based on any single work history might over or underestimate the true value. When there is high variability of single estimates about the true value, the study would have low reliability.

Unreliable exposure estimates present a different set of problems to epidemiological studies than do invalid estimates; moreover, the problems are not as serious and are more readily correctible. There is little published work that deals with reliability of self reported work histories. The few studies that have been done have dealt mainly with the reproducibility of portions of the work history—that is, questionnaire items on exposure, job titles, numbers of jobs, and job durations in selected occupational groups.6,7,8,9 They have not, however, focused on what impacts directly on epidemiological studies—namely, the reliability of complete work histories, and more specifically, cumulative exposure estimates derived from them. It is the objective of this study to do that and to relate the findings to the conduct of epidemiological studies.

Materials and methods
SOURCES OF DATA
This study, like its predecessor,1 arose from a 1976 survey by Fischbein et al10 that considered whether workers at a United States capacitor manufacturing facility developed abnormalities associated with PCBs.11 A repeat survey was conducted three years later. Work histories (self reported, interviewer assisted, dated chronological listings of recalled jobs) were obtained at both surveys for 164 subjects who were employed at the facility between June 1947 and March 1976, when PCBs were part of the production process. Job initiation and termination dates in these histories were usually specific to the month, but were designated as random month (1–12)/year if only year was given. Forty job

Institute of Environmental Medicine, Laboratory of Biostatistics and Epidemiology, New York University Medical Center, New York, NY 10010, USA
C R Rosenberg
categories, including inactive (illness, layoff, strike, etc) were distinguished with 20 of these comprising a main operation subgroup. These categories did not represent specific job skills but rather physically separated areas where individual steps of the production process occurred. Work history data, combined with average measurements of PCB concentrations in air gave two cumulative PCB exposure figures: one specific for June 1947 to December 1970 (early PCB period) when PCB mixtures with high chlorine content were used and another for January 1971 to March 1976 (late PCB period) when lower chlorinated mixtures were substituted. Such a distinction was necessary as published work had indicated that degree of chlorination is positively correlated with toxicity of PCBs. The presence of abnormalities was ascertained through physical examinations, blood testing, and medical history questionnaires. The finding of any significant associations between exposure to PCBs and abnormalities awaits further follow up.

**RELIABILITY CALCULATION AND DATA ANALYSIS**

Reliabilities of cumulative exposure to PCBs and duration of employment were obtained from one way variance component models. Specifically, for each variable, a random effects analysis of variance was performed where each of the 164 measurement pairs comprised the groups. The procedure produced estimates of within group variance (var(within)), and between group variance (var(between)), that were used to calculate an intraclass correlation coefficient, \( r_1 \) as:

\[
\frac{\text{var(between)}}{\text{var(between)} + \text{var(within)}}
\]

This intraclass correlation coefficient is an estimate of reliability under the assumption that any differences within the measurement pairs resulted from random variation about the true value. It has a maximum value of 1 when there is perfect agreement (no variability) within the measurement pairs and less than that when the paired measurements are not equal. Its minimum value is zero. In essence it is showing whether there is more similarity within groups than between groups. An approximate 95% confidence interval (95% CI) for \( r_1 \) can be calculated with the Fisher Z transformation. SAS was used for all statistical calculations.

An alternative method for estimating \( r_1 \) is from a validity study in which rough measures of a covariate (z) are compared with a “gold standard” (x). Here the reliability of z as a measure of x is equal to the square of the product moment correlation of x and z.

**Results**

Table 1 gives selected demographic and occupational variables for the study group. The age distribution was typical of a working population. Active duration, or actual time spent working at the production facility (excluding time lost to illness, layoffs, strikes etc), ranged from 25 to 347 months, with a mean of 179.5 months. Job change index, described in detail in the earlier paper of this series, was a function of number of job changes and duration within each job category and had a mean of 191.3. Its range was from 14 to 326. This corresponded to a mean of 14 and a range of two to 45 job changes per worker. The three variables were all roughly normally distributed. Most of the workers in the study group had completed 12 years of schooling (high school level) with a few having some college experience. The mean level was 11.3 years with a range of seven to 16 years. The distribution was skewed to the left.

Table 2 shows that cumulative exposure estimates for the early PCB period ranged from 0 to 2230 mg/m\(^3\) for the 1976 history and from 0 to 3738 mg/m\(^3\) according to the 1979 history. The respective means were 372.0 and 390.3 mg/m\(^3\). For the late PCB period, cumulative exposure estimates ranged from 0 to 1969 mg/m\(^3\) for both histories with respective means of 149.6 and 146.3 mg/m\(^3\). In both periods the difference between 1976 and 1979 survey means was not significant (paired t test), indicative of the fact that there were no systematic changes in cumulative exposure estimates from the 1976 to the 1979 survey. Indeed, any changes in exposure estimate from the initial to the second survey were equally likely to be upward or downward.

Table 3 gives estimates of the between and with-
in group variance components and the reliability coefficients for cumulative exposure to PCBs and duration of employment. Given the dependency of cumulative PCB exposure on a detailed work history and a three year gap between the surveys, the reliability coefficients for cumulative PCB exposure seemed surprisingly high. In the case of duration of employment, which was obtained as a response to a single question, the reliability was excellent. In all three cases the estimates were precise because of the large sample size for this analysis.

Table 4 shows the reliability coefficients for cumulative PCB exposure separately for workers above and below the overall median value for the job change index (195). It is clear for both the early and late PCB periods that reliability is lower for those with more job changes, and workers in the higher job change category produced coefficients about 8 to 15 percentage points lower than those in the lower change category. There was no evidence of any reliability differences due to age, sex, duration of employment, or educational level.

There were also indications in a subsample of workers with at least 20 years of service that time lapse had an influence on reliability—that is, the earlier portions of the work histories were less reliable than the more recent portions. The sample size for this analysis was, however, too small to establish any significant trends.

Discussion

Validity, which was considered in the initial paper in this series,1 is a measure of how close a measured variable is to the true value. Reliability, which is being examined here, is a measure of the consistency of repeated measures on a variable. Both of these measures are of great importance in assessing exposure in epidemiological studies. The implications of poor reliability, however, differ from those of poor validity.

It is of utmost importance in epidemiological studies that validity be sufficiently high to avoid serious misclassifications in assessment of exposure, which could lead to spurious findings. In other words, a lack of validity could lead to over or underestimation of the true effect measure (for example, relative risk) in a study depending on whether or not the resulting misclassification was random or non-random.19

Whereas it is also important to have highly reliable exposure data in a study, a less than desirable value is not necessarily a fatal flaw. The reason is that the effect of low reliability is to bias the effect measure in one direction only, toward the null value.19 Thus in studies that show a large effect low reliability is of no consequence—that is, the true effect is even larger than the estimated value. In studies where small effects are investigated, however, a lack of reliability could cause a real effect to evade detection.

To estimate reliability by repeat measurements, at least two measurements on the study subjects are required. When the underlying or true value of the reliability coefficient is above 0.5, taking two measurements per subject gives the optimal precision in estimating that coefficient. When the true value is below 0.5, additional measurements per subject are, however, needed to obtain adequate precision.20 The interval between measurements is also a consideration. If it is too small the study subjects may remember their original responses. If the interval is too large, systematic influences may intervene. It is not necessary to use the entire sam-

<table>
<thead>
<tr>
<th>Variance components:</th>
<th>Cumulative PCB exposure</th>
<th>Duration of employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early period</td>
<td>Late period</td>
</tr>
<tr>
<td></td>
<td>Low JCI</td>
<td>High JCI</td>
</tr>
<tr>
<td></td>
<td>Low JCI</td>
<td>High JCI</td>
</tr>
<tr>
<td>Between subjects</td>
<td>162203-0</td>
<td>198297-9</td>
</tr>
<tr>
<td>Within subjects</td>
<td>2733-2</td>
<td>22334-8</td>
</tr>
<tr>
<td>Reliability (rI)</td>
<td>0.983</td>
<td>0.899</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.977</td>
<td>0.864</td>
</tr>
<tr>
<td>Upper</td>
<td>0.987</td>
<td>0.925</td>
</tr>
</tbody>
</table>

Table 3 Overall reliability for study group

<table>
<thead>
<tr>
<th>Variance components:</th>
<th>Cumulative PCB exposure</th>
<th>Duration of employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early period</td>
<td>Late period</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between subjects</td>
<td>181881-59</td>
<td>36950-41</td>
</tr>
<tr>
<td>Within subjects</td>
<td>12534-00</td>
<td>1672-92</td>
</tr>
<tr>
<td>Reliability (rI)</td>
<td>0.936</td>
<td>0.957</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.914</td>
<td>0.941</td>
</tr>
<tr>
<td>Upper</td>
<td>0.952</td>
<td>0.968</td>
</tr>
</tbody>
</table>
people in the determination (which may add considerable expense) so long as the selected group is chosen randomly and is of adequate size to produce a precise estimate.

In the planning stages of a study the sample size necessary to detect a specified difference at a desired power may be adjusted upward to compensate for low reliability. Specifically, the required sample size would be calculated assuming no within-subject measurement variability and this value would be divided by the reliability coefficient to obtain the adjusted sample size.\(^{15-17}\) This assumes that the reliability is either known or can be determined from a pilot study. Alternatively, without increasing the sample size, the required number of multiple measurements per subject necessary to achieve adequate compensation may be calculated.\(^{19}\) In that case, the average of each subject's measurements would be used as that subject's exposure measure. This would require repeat measurements on the entire sample.

If it were only practical to obtain repeated measurements on a subsample of the study group, the resulting subsample variance components or reliability coefficients may still be used to correct effect measures in the overall analyses to the true values. Methods are available for correcting apparent correlation coefficients, regression coefficients, and relative risks.\(^{21}\) For epidemiological analyses that employ logistic regression, Rosner \textit{et al} \(^{22}\) provide a method for adjusting both the odds ratios and approximate CIs. If reliability is not determined before conducting a study it would be a good idea, if possible, to minimise any potential effects of low reproducibility by taking multiple measurements.

The high reliability found in this study group would seem to indicate that the self reported work histories were consistent in assessing cumulative exposure to PCBs, even after a three year gap. Given the detail of these work histories and the fact that repetition of them might be a costly or impractical undertaking, this consistency was reassuring (it may apply to other studies where self reported occupational information is used). The reliability coefficients found here, even in subjects with frequent job changes, compare favourably with those found in studies investigating the effects of diet on disease. Reliabilities of dietary information in such studies tend to fall in the 0.5–0.7 range.\(^{23}\)

Other occupational settings might, however, involve considerably more than the 40 job categories here, which would increase the potential for lower reliability. Furthermore, the reliability coefficients found in this study group may have overestimated the true comparability of the initial and repeat work histories. Of the 40 job categories, only 12 involved any exposure risk to PCBs—and of these 12, only half carried a moderate or high risk.

Thus a reasonable number of discrepancies between the subjects' work histories could have occurred involving no or low exposure categories that would not have produced inconsistencies in cumulative PCB exposure. Month by month comparisons of job categories for all the work history pairs did indeed show substantial proportions of inconsistency. On average, per subject, about 25% of the monthly intervals did not match for job category. The proportion of non-matching intervals for subjects with frequent job changes was substantially higher. Whereas some of this inconsistency can be accounted for by small chronological (frame shift) errors, many of the errors involved inconsistencies of categories or large time discrepancies within categories. The implication is that in occupational settings with higher proportions of exposed categories, more errors would involve the exposed categories and thus lead to lower reliability in cumulative exposure estimates.

The reliability coefficient for duration of employment was calculated for two reasons. Firstly, by contrasting it with the coefficient for cumulative exposures to PCBs, it showed that a directly reported variable would have a higher reliability. Secondly, and more importantly, it was to emphasise the fact that the reliabilities of not only the main variable of interest have to be considered in an epidemiological study, but also those of potential confounders. (Duration of employment was a confounding variable in several regression models examined in the first paper in this series.) The implications of low reliability of confounding variables are more serious in that measurement inconsistencies in them may bias the true effect either positively or negatively.\(^{24}\)

In closing, it should be mentioned that the concept of reliability applies not only to data that are derived from self reporting, as is the case in this paper and in dietary studies, but also to hard data measured by mechanical means. Assay derived data, such as prolactin concentrations in women at risk for breast cancer and cholesterol concentrations in subjects at risk of heart disease are also liable to vary within subjects and may thus have to be similarly assessed to determine if multiple measurements are required to improve reliability.\(^{25}\) It should also be noted that the estimates of cumulative PCB exposure evaluated in this study were derived not only from self reported work histories but from air measurements taken in the work areas, which were also subject to variation.

I thank Drs Roy Shore and Bernard Pasternack for reviewing this manuscript. This work was supported in part by PHS National Research Service Award (T32) ES07081 and center grant ES00928, both awarded by the National Institute of Environmental Health Sciences.
Requests for reprints to: Dr Carl R Rosenberg, New York University Medical Center, 341 East 25th Street, Room 206, New York, NY 10010, USA. Telephone 212–263–6500.


Accepted 9 November 1992

Destruction of manuscripts

From 1 July 1985 articles submitted for publication will not be returned. Authors whose papers are rejected will be advised of the decision and the manuscripts will be kept under security for three months to deal with any inquiries and then destroyed.
An analysis of the reliability of self reported work histories from a cohort of workers exposed to polychlorinated biphenyls.

C R Rosenberg

*Br J Ind Med* 1993 50: 822-826
doi: 10.1136/oem.50.9.822

Updated information and services can be found at: [http://oem.bmj.com/content/50/9/822](http://oem.bmj.com/content/50/9/822)

**Email alerting service**

*These include:*

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to: [http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to: [http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to: [http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)