

An analysis of the validity of self reported occupational histories using a cohort of workers exposed to PCBs

C R ROSENBERG,¹ M N MULVIHILL,² A FISCHBEIN,^{2,3} S BLUM⁴

From the Institute of Environmental Medicine,¹ Laboratory of Biostatistics and Epidemiology, New York University Medical Center, New York, NY 10016, Department of Community Medicine² and Division of Environmental and Occupational Medicine,³ Mount Sinai School of Medicine of the City University of New York, New York, NY 10029, and Office of Epidemiologic Surveillance and Statistics,⁴ New York City Department of Health, New York, NY 10013, USA

ABSTRACT An investigation was conducted to examine the validity of self reported work histories obtained from a sample of 326 capacitor manufacturing workers who had participated in an epidemiological study relating health abnormalities to exposure to polychlorinated biphenyls. Based on objective personnel records obtained for 288 members of the sample group, validity scores ranged from 20% to 100% with most falling between 70% and 75%. This validity range corresponded to misclassification proportions ranging from 13% to 29%. Exposure misclassification tended to be random for men but non-random for women (overestimation of exposure). By means of multiple regression analysis, the diversity of the job categorical pattern was found to be a significant independent predictor of validity. The factors sex, duration, and time lapse were also significant validity predictors but only interactively with job diversity. There were also indications that interviewer skill could be an important factor. Knowledge of the magnitude of these factor effects could be critical in planning retrospective epidemiological studies.

Obtaining accurate exposure data in retrospective epidemiological studies can be difficult, especially when the study subjects are the sole source of the data. Inaccurate self reported data can lead to exposure misclassification, resulting in null biased, undetected, or spurious associations. The dependence of many epidemiological studies on self reported data initially led researchers to test the validity of single discrete exposures,¹⁻³ chronic disease reports,⁴⁻⁷ and self administered questionnaires.⁸⁻¹⁰ Subsequent research included similar type data sets but used analytical techniques to investigate causes of inaccurate reporting.¹¹⁻¹⁵ Validity testing and analysis of work histories used to quantify harmful occupational exposures should have soon followed but was not seriously considered until recently when Baumgarten *et al* assessed self reported work history

accuracy and factors that influenced it in the context of a case-control study of occupation and cancer.¹⁶ To advance knowledge further in this area, we compared corresponding self reported and company compiled job records, showing how self reporting errors can lead to exposure misclassification of study subjects and how multiple regression can be used to analyse and quantify the effects of various factors on work history validity. Unlike the studies cited above, which with one exception⁸ expressed validity as a group function, we have measured and analysed validity on an individual basis.

Materials and methods

SOURCES OF DATA

This study arose from a 1976 survey by Fischbein *et al*,¹⁷ the aim of which was to determine if workers at a US capacitor manufacturing plant, exposed to polychlorinated biphenyls (PCBs), developed clinical abnormalities associated with PCBs.¹⁸ Work histor-

ies (self reported, interviewer assisted, dated chronological listings of recalled jobs) were obtained for 326 individuals who were employed 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 by us as "random month (1-12)/year" if only a year appeared. Forty job categories 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. Often unreported, but designated by us as a job category, were non-working (illness, layoff, strike, etc) intervals. The work history data, combined with average PCB air level measurements, yielded 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.¹⁹ Such a distinction was necessary since reports have indicated that degree of chlorination is positively correlated with PCB toxicity.¹⁸ The presence of abnormalities was ascertained through physical examination, recording of medical history, and blood testing. The finding of any significant associations between PCB exposure and abnormalities awaits further follow up.

Our acquisition of microfilmed personnel records compiled by the capacitor production company (a presumed "truth criterion" against which the self reported work histories could be compared) made the proposed validity analysis possible. Corresponding personnel records were available for 288 of the work histories. The personnel records were more comprehensive than the work histories, not only listing the chronology of assigned jobs but time lost to illness, layoffs, strikes, suspensions, military service, or pregnancy. Dates of job initiation and termination were day specific, permitting precise tenure determinations. All jobs were listed in code with titles being ascertained from a company master list. This list was translated into the categorical system used by Fischbein *et al.*,¹⁷ facilitating the calculation of company specific early and late period PCB exposure figures and our validity determination. Demographic information provided included date of birth, sex, height, weight, race, and educational level. All the data were recorded at the time of occurrence, uninfluenced by recall or interviewer bias, and thus represented a relatively objective base of validity. The likelihood of systematic biases related to PCB exposure in these records was negligible since they were compiled for bookkeeping purposes. Any bookkeeping errors that might have occurred, however, could not be checked.

VALIDITY, MISCLASSIFICATION, AND JOB DIVERSITY INDEX

For each subject, the employment tenure (hire date to March 1976) was broken down into one month intervals. A particular interval—for example, August–September 1959—was designated as valid when the self reported and company listed job categories matched within this period. The number of valid intervals divided by the employment tenure (in months)—that is, the per cent agreement between a work history and personnel record—was defined as the validity score. Full tenure validity scores were obtained for the entire sample. For subjects with 23 or more years of tenure, subperiod scores were also calculated: for 1950/9, 1960/9, and 1970/6. The validity score was sensitive to omission, insertion, overreport and underreport, and chronological (frame shift) errors. It was necessary to take the latter into account because of changes over time in plant exposure conditions. Changing conditions imply that job specific exposure levels are functions of time and thus the same job category at different calendar periods may represent a different exposure. Any portion of these scores accounted for by chance agreement would have been minimal because of the large number of possible job categories so that a correction for this effect was deemed unnecessary.

The proportion of study subjects misclassified within the early and late calendar periods was obtained by comparing, within each period, corresponding self reported and company derived cumulative exposure categories. Four categories of exposure were possible for early and late calendar periods, derived from quartiles of the respective distributions of company derived cumulative exposure figures. Misclassification was defined as the placement of a subject by a self reported history in an exposure category other than that determined by the company. Directionality of misclassification was noted as either upward or downward relative to the personnel record category.

It was understood when planning this study that frequent job changes, which produce complex work patterns, would limit accuracy. Nevertheless, it was desirable before considering this effect to create a weighted job change index that gave lesser importance to brief job tenures (which only marginally influenced complexity and cumulative exposure) and standardised job change frequency for subjects with varying employment tenures. These requirements were met by the Shannon-Wiener index²⁰ which was designed to measure diversity (complexity) in ecological communities. In the present epidemiological application it was designated job diversity index, measuring the complexity of a subject's personnel record work pattern. Com-

putation was as follows:

$$\text{Job diversity index} = - \sum_{i=1}^k (n_i/N) \ln(n_i/N) \times 100$$

where N = employment tenure (from hire date to March 1976) in months

n_i = tenure within job category i in months

k = total number of job categories.

In calculating job diversity, each category change denoted a new job tenure whether or not that particular category appeared earlier in the work pattern. The minimum possible value is zero; for a single job category spanning an entire employment tenure. The highest values would occur in cases of given employment, tenures being divided into increasingly numerous job categories.

ANALYSIS

SAS was used for all parametric statistical analyses in this study.²¹ Potential predictors of validity and crude validity scores were first descriptively examined. Then multiple linear regression, through simultaneous adjustment, allowed all (measurable) factor effects (including possible interactions) on validity to be assessed in an unconfounded manner. To maximise predictive ability, factors measured on a continuous scale were maintained in that format. Factors that could not be assessed through regression (interviewer skill, recall ability, cooperativity, and stress) were examined qualitatively. Non-randomness of categorical misclassification was assessed non-parametrically with the sign test.

Results

EXAMINATION OF POTENTIAL VALIDITY PREDICTING FACTORS

Descriptive statistics for the factors age (at interview), (active) duration, and job diversity are shown in the table. These were potential validity predictors that varied considerably in the study group. Race and educational level also had validity predicting potential but were excluded for not meeting the variability criterion. Each factor was segregated by sex, another potential validity predictor.

The age distributions for men and women were normal. The ranges were identical but the respective means, 41.7 and 47.3 years, differed significantly (Student's t test, $p < 0.0001$). (Active) duration, or actual time spent working at the production plant (excluding time lost to illness, layoffs, or strikes, for example) was normally distributed and ranged from 21 to 352 months. The difference between the male and female means, 171 and 179 respectively, was not significant. Job diversity was also normally distributed and ranged from 14 to 326. This corresponded to a range of two to 45 job changes per individual. The mean female diversity, 214, was significantly higher than that for men, 170 (Student's t test, $p < 0.0001$), indicating that women had considerably more complex work patterns. The mean number of job changes was 17 for women and 11 for men.

VALIDITY PROFILE OF THE STUDY GROUP

The distributions of the male and female validity scores are shown in fig 1. The ranges of scores illustrate the deviations of the self reported work histories from the actual work patterns. Male validity scores ranged from 20% to 99% (mean = 76.6) whereas those for women fell between 15% and 100% (mean = 74.3). Note the skewness of the distributions with the majorities of scores located between 75% and 100%.

EVIDENCE OF MISCLASSIFICATION

Figure 2 (left half) illustrates the degree to which study subjects were correctly classified as to cumulative PCB exposure. In each (sex and exposure period specific) 4×4 cross classification subtable consistency of the work history and personnel record categories (1 = lowest exposure quartile) may be observed. The shaded diagonals represent correct classification, whereas the regions above and below indicate upward and downward misclassification, respectively. For the early period exposure 12.6% of the women and 21.5% of the men were misclassified. The respective figures for the late period were 18.6% and 13.3%. There was no consistent sex difference as a greater proportion of men was misclassified for early period exposure whereas the opposite was true

Selected demographic (Std dev) and occupational factors for members of the study group

Factor	Sex	No	Mean	Std dev	Range
Age at interview (years)	M	145	41.7*	12.3	22-65
	F	143	47.3	10.3	22-65
Active duration (months)	M	145	170.5	92.1	21-347
	F	143	179.1	84.9	25-352
Job diversity index	M	145	170.4*	51.2	26-288
	F	143	214.4	57.4	14-326

*Male female difference is significant (Student's t test, $p < 0.0001$).

for the late period. When misclassified, however, women tended to shift to a higher exposure category. This tendency was significant for both the early and late periods (sign test: $p = 0.004$ and 0.003 , respectively). In contrast, male misclassification was random. Comparing early and late period misclassification proportions was unwarranted since late period quartiles were considerably broader than those for the early period.

On the right half of fig 2 results are listed for a repeat survey done by Fischbein in 1979 involving a subset of the original group and covering the same exposure period, 1947-76. With 45 additional months of recall, consistently higher misclassification proportions were evident. From a range of 12.6% to 21.5% in 1976 the range was from 16.5% to 29.1% in 1979. Within this reassessed subgroup the tendency for greater misclassification in 1979 was significant (McNemar's test: $p = 0.005$).

FACTORS INFLUENCING VALIDITY

In a regression analysis using the entire sample ($n = 288$) the effects on overall (full tenure) validity of age, duration, job diversity, and sex were evaluated with maximum power. The assessment of time lapse—that is, the validity variation over early (1950-9), middle (1960-9), and late (1970-6) subperiods—required a

sample with sufficient tenure (at least 23 years) to capture enough early phase years and thus restricted analysis to 127 eligible subjects. Job diversity, age, and sex were included in the time lapse analysis but not duration, since this latter factor was, by definition, homogeneous in this sample. Linear models including main effects and appropriate second and third degree interactions were fitted to the data after the transformation: $\text{validity}^* = \sin^{-1}(\text{validity}/100)^{\ddagger}$, which is appropriate for normalising variables recorded as percentages.²² The inclusion of any other higher order terms was not appropriate. Backward elimination guided by F tests comparing full and reduced models yielded optimal validity predicting models.

The full sample analysis showed job diversity to be the only significant independent predictor of validity. Sex and duration were also significant, but only interactively with diversity—that is, statistically influencing validity only when diversity exceeded zero and in practical terms only above a critical diversity value. As fig 3 shows, validity (reconverted to per cent from radians) declined with increasing job diversity. At zero diversity (a single job spanning an entire employment tenure) predicted validity was logically 100%. This validity level persisted until a diversity of 40, above which it descended toward lower ranges, but at a slower rate for women. The rate of validity decline attributable to diversity (slope of female regression line) was $2.8 \pm 0.4 (\times 10^{-3}) \uparrow$ rad/div unit. (Male) sex added 0.4 ± 0.2 rad/div unit to this decline rate yielding a male regression line slope of 3.2 rad/div unit. At highest job diversity levels in this study group, predicted female and male validity scores were as low as 50% and 60%, respectively. A female advantage (equal or better scores at any given diversity) was not apparent from the mean validity scores in the "validity profile" section, which were unadjusted for confounding due to the male-female job diversity difference. Regression analysis removed the effect of this confounder and brought out the underlying male-female validity difference.

The effect of duration was to weaken, interactively, the validity reducing effect of diversity. Each additional month of duration attenuated the decline rate by 0.003 ± 0.001 rad/div unit. Figure 3, for clarity, only shows curves for 170 months, the mean study group duration level, but to visualise the duration effect, longer tenures would have yielded curves with slower rates of descent and vice versa. Age was eliminated as a validity predicting factor since it was so highly collinear with duration. This was of no major

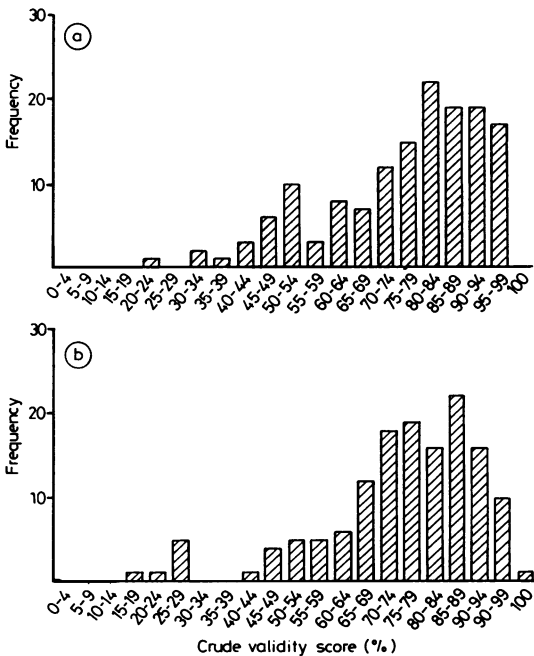


Fig 1 Frequency distribution of unadjusted validity scores for (a) men and (b) women.

*Transformed (to radians: $1.57 \text{ rad} = 100\%$).

† ± Following regression coefficient is 95% confidence interval. All coefficients are multiplied by 10^{-3} radians/diversity unit.

1979 Survey

Early calendar period (1947-70)

Women					Men				
EXCAT (WH)*					EXCAT (WH)				
1	2	3	4	T	1	2	3	4	T
1	6	4	0	0	1	13	3	1	0
2	2	13	4	0	2	1	13	3	1
3	1	0	10	6	3	1	4	14	7
4	0	0	1	22	4	0	0	2	16
T	19	17	15	28	T	15	20	20	24
EXCAT (PR) †					EXCAT (PR)				
1	16	4	0	0	1	13	3	1	0
2	2	13	4	0	2	1	13	3	1
3	1	0	10	6	3	1	4	14	7
4	0	0	1	22	4	0	0	2	16
T	19	17	15	28	T	15	20	20	24

Misclassified upward: 14§
 downward: 4
 Prop misclassified: 18/79=22.8%

Misclassified upward: 15
 downward: 8
 Prop misclassified: 23/79=29.1%

1976 Survey

Early calendar period (1947-70)

Women					Men				
EXCAT (WH)*					EXCAT (WH)				
1	2	3	4	T	1	2	3	4	T
1	7	6	0	0	1	25	3	1	0
2	1	30	4	0	2	4	23	6	0
3	1	1	25	4	3	0	8	25	5
4	0	0	0	36	4	0	0	2	33
T	29	37	29	40	T	29	34	34	38

Misclassified upward: 14§
 downward: 3
 Prop misclassified: 17/135=12.6%

Misclassified upward: 15
 downward: 14
 Prop misclassified: 29/135=21.5%

Late calendar period (1971-76)

Women					Men				
EXCAT (WH)					EXCAT (WH)				
1	2	3	4	T	1	2	3	4	T
1	18	3	0	0	1	44	1	1	2
2	4	34	6	0	2	2	23	1	1
3	1	1	23	11	3	0	2	29	3
4	0	0	0	39	4	0	0	6	28
T	23	38	29	50	T	46	26	37	34

Misclassified upward: 20§
 downward: 6
 Prop misclassified: 26/140=18.6%

Misclassified upward: 9
 downward: 10
 Prop misclassified: 19/143=13.3%

Late calendar period (1971-76)

Women					Men				
EXCAT (WH)					EXCAT (WH)				
1	2	3	4	T	1	2	3	4	T
1	8	1	2	0	1	26	1	1	1
2	3	22	3	1	2	0	9	3	0
3	2	1	12	6	3	0	2	19	1
4	0	0	2	22	4	2	0	2	12
T	13	24	19	29	T	28	12	25	14

Misclassified upward: 13
 downward: 8
 Prop misclassified: 21/85=24.7%

Misclassified upward: 7
 downward: 6
 Prop misclassified: 13/79=16.5%

Fig 2 Misclassification results for 1976 and 1979 surveys. *Exposure category according to work history; †exposure category according to personnel record; ‡shaded diagonals represent correctly classified subjects (above diagonal: misclassified upward, below diagonal: misclassified downward); §tendency for upward misclassification was significant (sign test: $p < 0.01$).



Fig 3 Regression lines describing relation between adjusted validity and factors of job diversity, sex, and duration (full sample).

concern, though, since some validity studies^{4,6} have shown age to be a poor predictor of recall ability in normal working age individuals.

The R^2 value, or fraction of variability explained by the chosen model, was 0.39. If the effects of recall ability, interviewer skill, cooperativity, and stress could have been quantified, however, the amount of explainable variability would have been greater.

In the time lapse analysis job diversity, again, was the only significant independent predictor of validity. As fig 4 shows, validity scores generally declined with increasing job diversity. Nevertheless, time lapse was borne out by this regression to be a significant interactive determinant of the rate at which this decline occurred. Despite the significance of this effect, it was conditional, affecting validity only above a critical diversity value. Logically, time lapse would not be expected to hinder the self report accuracy of a worker who remained in a single job category for an entire employment tenure. Predicted validity at zero diversity was 100% and remained so until a diversity of 40, when time lapse became operative. The decline rate for 1970-6 subperiod (slope of 1970s regression line) was 1.9 ± 0.5 rad/div unit, which was barely perceptible in fig 4, with validity remaining above 90% over almost the entire diversity range. This rate applied to both men and women as the effect of sex was not statistically significant in this subperiod.

In the 1960-9 subperiod the decline rate not only became evident but was also significantly modified by sex. Decline intensification attributable to this subperiod was 0.6 ± 0.3 rad/div unit, whereas that due to (male) sex was 0.7 ± 0.4 rad/div unit, leading to female and male decline rates (1960s regression line slopes) of 2.5 and 3.2 rad/div unit, respectively. At the highest study group diversity levels, predicted validity

in this subperiod dropped to about 70% for women and 50% for men. The 1950-9 subperiod introduced a decline rate effect of 1.7 ± 0.3 rad/div unit. The contribution of sex was again 0.7 ± 0.4 rad/div unit because the magnitude of this latter effect did not differ significantly between the 1950-9 and 1960-9 subperiods. The female regression line slope (validity decline rate) was 3.6 rad/div unit, a near doubling of the 1970s rate. The combination of maximal time lapse and (male) sex produced the ultimate decline rate of 4.3 rad/div unit. Predicted validity scores at maximal diversity were in the 20-40% range.

These results indicate equal or better female validity at equivalent diversity and subperiod. It was also apparent that the time lapse effect was dominant over that of sex in that at least six years of the former was required to produce any male-female difference. The R^2 for this regression was 0.43, higher than in the full sample analysis.

INTERVIEWER EFFECTS

There were suggestions through further regression analysis that interviewer skill affected validity scores of study subjects. Negative interviewer effects on validity might have been due to careless date recording, inaccurate coding, and the fact that non-working tenures were rarely picked up unless they exceeded six months. A four month strike (October 1968 to February 1969), for example, appeared on only one of 288 work histories. Unfortunately, though, significant differences between interviewers could not be confirmed owing to small sample sizes per interviewer.

THEORETICAL EXPECTATIONS

In fig 5 the aim is to show qualitatively how validity score distributions and misclassification proportions in any occupational setting may be predicted from knowledge of various factor levels. Understanding



Fig 4 Regression lines describing relation between adjusted validity and factors of job diversity, time lapse, and sex (time lapse analysis subgroup).

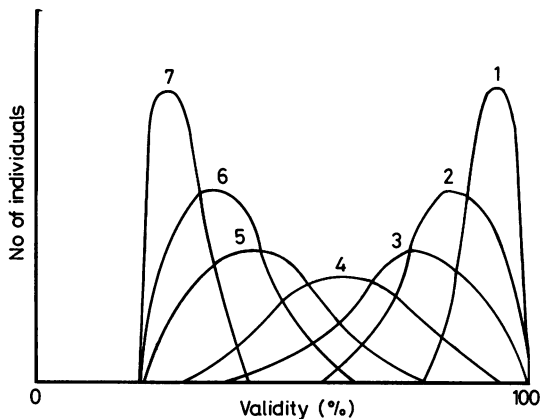


Fig 5 Theoretical validity distributions for selected job diversity ranges (1: 0-79, 2: 80-159; 3: 160-239, 4: 240-319, 5: 320-399, 6: 400-479, 7: > 479).

the major controlling role of job diversity, potential validity score distributions were linked to a progression of diversity ranges. Distribution 1 was tied to the diversity range 0-79, 2 to the range 80-159, and so on. The linkage of a specific distribution to a particular diversity range is not rigid and may be influenced by the interactive effects of time lapse, duration, sex, cooperativity, stress, or interviewer skill. A normal distribution of recall abilities is assumed in overcoming work pattern complexity.

Occupational groups with simple work patterns would produce distributions similar to 1-3 whose skewed appearance, despite an assumed normal distribution of recall abilities, would be due to the constraint of 100% validity. Neither good recall ability nor favourable factor status would be critical for accuracy and misclassification proportions would be minimal. Higher diversity would yield a distribution similar to 4 which is no longer constrained by the 100% validity level, yielding a normal appearing distribution of validity scores. Interactive factor influences and the potential for misclassification would be greater. Still higher diversity ranges would produce distributions similar to 5-7, where subjects, regardless of recall ability, would be at an accuracy disadvantage. Interactive factor effects would also assume greatest importance. Constraint imposed by the minimum validity barrier accounts for the skewed appearance of distributions 5-7. Misclassifications under these circumstances might be overwhelming, making effective studies difficult if not impossible. Note that the minimum score level is depicted as greater than zero, because it is unlikely that high diversity alone could produce zero validity. If it did occur it would take unusually poor interviewing

technique, stress, or low cooperativity to produce it. The range of job diversity in this study group, 14-326, would probably encompass diversity ranges 1-4 (see fig 1).

Discussion

In epidemiological studies, whenever possible, objective sources of exposure information should be used. Increasingly, employers and unions, with the cooperation of government and medical personnel, are making this possible by compiling daily work records in industries where harmful exposures may occur. Even so, there will continue to be cases where objective sources will not be available or, if obtainable, unsatisfactory, owing to incompleteness, lack of specificity, or subjective recording. The only alternative, then, would be to use self reported information to assess exposure.

In this study group there were indications that while self reporting was certainly not a haphazard procedure, it may have fallen short of what was necessary to ensure dependable relative risk estimates. Mean non-validity (100% minus mean validity) of just 25% led to exposure misclassification proportions well above levels known to cause substantially null biased risk estimates. Breslow and Day showed how a misclassification proportion as low as 10% could reduce a relative risk estimate by more than 50% from the actual value.²³ Also, the significant upward exposure misclassification of women found here could have led to spurious associations between abnormalities and PCB exposure if the abnormalities influenced the upward shift. Why only women tended to overestimate time spent in high exposure categories is unclear. Findings from the 1979 repeat survey pointed out a further problem in conducting self reporting studies in that any delay in collecting exposure data could further contribute to misclassification.

An understanding of why the validity level in this study group may have been insufficient to prevent serious misclassifications may be important for future self reporting studies in terms of their feasibility, scope and limitations, and freedom from preventable inaccuracies. The regression analyses, in elucidating the roles and interrelationships of various factors in predicting validity, attempted to provide such an understanding.

With respect to study feasibility, knowledge of job diversity beforehand may be advantageous. If diversity is low (< 80), recall will be easy (uninfluenced by interactive factors), misclassification will be minimal, and feasibility will not be an issue. Increasingly higher diversity would make recall progressively more difficult (and more prone to complication by second-

ary factors), leading to greater misclassification or possibly calling feasibility into question. In this study group 40 different job possibilities resulted in a mean job diversity of 192 and misclassification proportions ranging up to 29%. This implied that in industries with a large number of job types it may be difficult to obtain valid self reported information about occupational exposures.

Time lapse may also have important implications for study feasibility in that: (1) studies investigating diseases with long latent periods would depend on exposures dating back 20 or more years requiring focus on less valid early phases of work histories and (2) more serious exposures tended to occur in earlier years due to lack of knowledge about effects and fewer regulatory protections, also requiring concentration on early phases of work histories. The latter was certainly a factor in this study group as most exposures to higher levels of more toxic PCBs occurred during the 1950-9 subperiod. (If the survey by Fischbein *et al*¹⁷ had dealt only with exposure during this subperiod, misclassification proportions would have been much higher than the full tenure figures reported here.)

If a study is to be done scope and limitations should be considered. For example, high diversity may exist in a potential study group but be confined to only a few subjects or one subgroup. In this event it might be wise to eliminate such individuals, but weighing this against the loss of power. In this study group elimination of subjects with diversities exceeding 250 could have possibly reduced the misclassification proportions to acceptable levels. The time lapse factor would also come into play in that studies examining longer periods of exposure history would present more difficulties than ones covering shorter periods. In the case of a long exposure history analysis may have to be restricted to more recent exposures.

Interviewer error, stress, low cooperativity, and poor recall, although unmeasurable by regression, were probably partially responsible for misclassifications in this study group. The critical point is that these factors are to some degree under the control of the investigator and thus inaccuracies due to them are preventable. To minimise or eliminate their effects: (1) trained interviewers should be employed, and in optimum numbers to maximise information gathering (help subjects compensate for high diversity) while minimising bias, (2) subjects should be asked to prepare in advance for the reporting session, and (3) the interviewing environment should also encourage optimal reporting by assuring low stress and high cooperativity.

Of less practical importance (at least at this stage) but still of academic interest were the significant interactive effects of duration and sex. The positive cor-

relation of duration with validity may have been due to the fact that it distinguished between workers with long and short tenures who differed in terms of recall ability, stress level, or cooperativity. Perhaps workers with long tenures had higher validity due to greater identification and familiarity with plant operations. The significant female validity advantage was in agreement with other studies of validity that have found women to report more accurately than men. An explanation, as in the case of duration, could male female differences in one or more of the unmeasurable factors. It would be interesting to see if this sex effect is typical of self reporting.

In conclusion, we have found that validity of self reported work histories may be quantified in a useful way by job diversity and have found various determinants of validity. Further research is recommended on other occupational groups with various diversity ranges, factor levels, exposures (especially neurotoxic considering memory effects), and population types. The methodology used in this study might also be extended to evaluate validity in other situations where self reported data are required.

We thank Drs I J Selikoff, J Thornton, A M Langer, B S Pasternack, M Marmor, and R E Shore for help in preparing 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 C R Rosenberg, New York University Medical Center, 341 East 25th Street, Room 206, New York, NY 10010.

References

- 1 Dunn JE, Jr, Buell P. Association of cervical cancer with circumcision of sexual partner. *J Natl Cancer Inst* 1959;108:749-64.
- 2 Lerman SJ, Lerman LM, Nankervis GA, Gold E. Accuracy of rubella history. *Ann Intern Med* 1971;74:97-8.
- 3 Goebel WM. Reliability of the medical history in identifying patients likely to place dentists at an increased hepatitis risk. *J Am Dent Assoc* 1979;98:907-13.
- 4 Madow WG. *Interview data on chronic conditions compared with information derived from medical records.* (National Center for Health statistics report ser 2, No 23.) Washington: US DHEW, Public Health Service, 1967.
- 5 Commission on Chronic Illness. *Chronic illness in the United States.* Vol IV. *Chronic illness in a large city—the Baltimore study.* Cambridge: Harvard University Press, 1957.
- 6 US National Health Survey. *Health interview responses compared with medical records.* (Health statistics ser D, No 5.) Washington: US DHEW, Public Health Service, 1961.
- 7 Meltzer JW, Hochstim JR. Reliability and validity of survey data on physical health. *Public Health Rep* 1970;85:1075-86.
- 8 Pecoraro RE, Inui TS, Chen MS, Plorde DK, Heller JL. Validity and reliability of a self-administered health history questionnaire. *Public Health Rep* 1979;94:231-8.
- 9 Sacks JJ, Krushat WM, Newman J. Reliability of the health haz-

- ard appraisal. *Am J Public Health* 1980;**70**:730-2.
- 10 Petitti DB, Friedman GD, Kahn W. Accuracy of information on smoking habits provided on self-administered research questionnaires. *Am J Public Health* 1981;**71**:308-11.
 - 11 Chamberlain G, Johnstone FD. Reliability of the history. *Lancet* 1975;i:103.
 - 12 Norell SE. Accuracy of patient interviews and estimates by clinical staff in determining medication compliance. *Soc Sci Med* 1981;**15E**:57-61.
 - 13 Brady WF, Martinoff JT. Validity of health history data collected from dental patients and patient perception of health status. *J Am Dent Assoc* 1980;**101**:642-5.
 - 14 Corwin RG, Krober M, Roth HP. Patients' accuracy in reporting their past medical history: a study of 90 patients with peptic ulcer. *J Chronic Dis* 1971;**23**:875-9.
 - 15 Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982;**116**:114-22.
 - 16 Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiological purposes. *Am J Epidemiol* 1983;**118**:583-91.
 - 17 Fischbein A, Wolff MS, Lilis R, Thornton J, Selikoff IJ. Clinical findings among PCB-exposed capacitor-manufacturing workers. *Ann NY Acad Sci* 1979;**320**:703-15.
 - 18 Kimbrough RD, ed. *Halogenated biphenyls, terphenyls, naphthalenes, dibenzodioxins, and related compounds*. Amsterdam: Elsevier, 1980.
 - 19 Jones M. *Industrial hygiene survey of the two capacitor-manufacturing plants*. Cincinnati: National Institute for Occupational Safety and Health, 1978.
 - 20 Risser PG, Rice EL. Diversity in tree species in Oklahoma upland forests. *Ecology* 1971;**52**:876-80.
 - 21 SAS Institute. *SAS user's guide*. Cary NC: Statistical Analysis Systems Inc, 1979.
 - 22 Pasternack BS, Shore RE. Analysis of dichotomous response data from toxicological experiments involving stable laboratory mouse populations. *Biometrics* 1982;**38**:1057-67.
 - 23 Breslow NE, Day NE. *Statistical methods in cancer research. Vol 1. The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980:114. (Sci publ No 32.)