International Journal of Epidemiology
© International Epidemiological Association 1996

Vol. 25, No. 4
Printed in Great Britain

The Repeatability of Self-Reported Exposure after Miscarriage

A FARROW,* S C FARROW,** R LITTLE,† J GOLDING* AND THE ALSPAC STUDY TEAM

Farrow A (Institute of Child Health, University of Bristol, 24 Tyndall Ave, Bristol, Avon, UK), Farrow S C, Little R, Golding J and the ALSPAC Study Team. The repeatability of self-reported exposure after miscarriage. *International Journal of Epidemiology* 1996; **25:** 797–806.

Background. The Avon Longitudinal Study of Pregnancy and Childhood is a prospective study of women who were resident in Avon and who were expected to deliver a baby between April 1991 and December 1992.

Methods. The study provided an opportunity to test the repeatability of responses from 220 women who experienced a miscarriage and who reported exposure to occupational substances and common household products and appliances in two questionnaires. The first questionnaire was completed in the early part of the pregnancy and the second after the miscarriage. Women were asked to score their frequency of exposure on a five-point scale from 'daily' to 'never'. Their responses were analysed to assess the degree of agreement between replies to identical questions in the two questionnaires using the kappa statistic. A new frequency variable was created which compared the replies for the two questionnaires; this was analysed for all exposures by cross-tabulation with possible explanatory variables (age of mother, social class, history of miscarriage and the time lag between questionnaires).

Results. In general there was good agreement in the reported exposures to 48 substances and products. The results showed a small and consistent pattern of reporting exposures less frequently in the second questionnaire, i.e. after miscarriage. This was not explained by the analysis of possible confounding variables. Given the literature, the authors had expected to find a shift in the opposite direction.

Conclusion. The study reinforces the need to be cautious when using the results from single surveys of retrospective self-reported exposure.

Keywords: exposure, self-report, repeatability, miscarriage, bias, questionnaire

Repeatability or reliability can refer to the agreement of information and can be assessed when a measurement is performed twice or the same question asked twice. When the results are different this may express variability in the measurements rather than real change in the variable being measured. Indices to express the results of repeatability or reliability tests in statistical terms include simple reliability coefficients or per cent agreements. An index that makes allowance for the contribution of chance agreement is kappa[1–3] and there are different conventions for interpreting the values.[4,5]

Studies of reliability include that for passive smoke exposure in a study of lung cancer using re-interviews for a sample of non-smoking cases and controls. Agreement rates were generally higher for controls. They were also higher when considering whether the parents or spouses had ever smoked, with lower agreement rates for duration or intensity of smoking i.e. higher agreement rates with narrower exposure categories.[6] A comparison of prospective and retrospective assessments of diet suggests biased associations between fat intake and risk of breast cancer.[7] Few have assessed information obtained before and after miscarriage or in early and late pregnancy.

This paper reports on a study based on data from a prospective cohort study of pregnant women who had completed a questionnaire during the first trimester of their pregnancy (first questionnaire). Women who subsequently miscarried and accepted an invitation to take part in a further study were asked to complete another questionnaire (post miscarriage questionnaire), which included identical questions to those in their first questionnaire. The repeated responses offered an opportunity to examine agreement. The objectives of this study were to:

i) determine the repeatability of self-report for frequency of use of specific common household products and appliances and agents encountered in either work or leisure activities after the adverse event of a miscarriage.

ii) determine any differences in the repeatability in relation to age, socioeconomic status, previous miscarriage and the time lag between answering the two

* Institute of Child Health, University of Bristol, 24 Tyndall Ave, Bristol, Avon, UK.
** Health Research Centre, University of Middlesex, UK.
† National Institute of Environmental Health Sciences, NC, USA.

questionnaires. All these data were available on the study women.

SUBJECTS AND METHODS

The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) is a prospective cohort study of women who were resident in Avon and expected to deliver a baby between 1 April 1991 and 31 December 1992. The total number enrolled in the study was 14 893 which was a participation rate of 85% of eligible women. Of the total enrolled, 717 pregnancies did not extend beyond the end of the 23rd week of gestation; these included the miscarriages.[8] The post miscarriage questionnaire was sent to 561 of the 717 women. In some cases the midwives considered it inappropriate to send the questionnaire; in other cases the women had moved from the area. The post miscarriage questionnaire was completed by 467 women, yielding a response rate of 83% of available participants. Some of these women had enrolled early enough in their pregnancy to have completed a first questionnaire before the failure of the pregnancy (n = 287) and of these, 222 were miscarriages. The remaining 65 included women who had experienced a fetal death, termination of pregnancy for medical reasons or other adverse outcome of pregnancy (e.g. blighted ovum) that was not defined as a miscarriage. These final 222 women became the study group for the test of repeatability to responses. The first questionnaire, sent out in early pregnancy, included questions on the home and social environment such as use of household chemical products, medication, electrical equipment, work or hobby agents and job history since age 16.

As soon as notification of the pregnancy loss was received at the study centre, a letter of condolence was sent to the woman with an invitation to take part in a special study investigating their loss. If the invitation was accepted, the woman was sent a post miscarriage questionnaire concerned with environmental aspects surrounding early pregnancy. The post miscarriage questionnaire contained the same list of environmental variables to that found in the first questionnaire. The layout and the order of questions was identical and the questionnaires were of similar length. Both questionnaires enquired about exposures in the first trimester of the study pregnancy. Frequency of exposure was reported on a five-point scale, from daily through never. This investigation is concerned with those questions that were identical in both questionnaires and the comparison of the answers for frequency of exposure from each. These comparisons were defined as 'less frequent' 'same frequency' or 'more frequent'. When the post miscarriage questionnaire responses differed from the first questionnaire it was possible to assign them to 'less frequent' or 'more frequent'. For example, if the response changed from 'daily/most days' in the first questionnaire to 'less than once a week' in the post miscarriage questionnaire, it was classified as 'less frequent'.

Statistical Analysis

Measures of agreement between the paired responses were assessed using the kappa statistic. These were based on the four categories of frequency of response. Two of the frequency categories 'daily' and 'most days' were combined into a single category because of small numbers. Where there were zeros in frequency categories in one questionnaire and not in the other, further combination of contiguous groups was necessary to calculate kappa values.

Subtraction of the value of the response in the post miscarriage questionnaire from the value in the first questionnaire resulted in three new categories for the exposure variables: 'same' frequency of use, 'less' frequent use or 'more' frequent use. The statistical significance of these variables was assessed using the Wilcoxon non-parametric signed rank test. This was calculated on the 'less frequent', 'same' and 'more frequent' components between the two sets of responses. This tests the hypothesis that there are no differences between two paired populations of ordered-metric scores. Although a 'significance' level of 0.05 was used for the presentation of tables, in view of the multiple testing of significance, levels of 0.01 are used for the purposes of discussion.

Analysis of possible explanatory variables included maternal age categorized into three groups, <24, 25–34 and ≥ 35 years, social class of the mother (based on her most recent job) and banded into manual and non-manual classes, number of previous miscarriages categorized into none or one or more and time lag between questionnaires grouped into <39 days, 40–65 days and >65 days. For each of these variables χ^2 analysis was performed on the cross-tabulation against the frequency variable in the three categories 'less frequent', 'same frequency' or 'more frequent'.

RESULTS

Of the 222 women studied, 17 women gave no information on exposure in the first questionnaire. A further 21 gave incomplete answers to some questions in one or other of the questionnaires. Table 1 shows the overall percentages of women reporting frequency of

TABLE 1 *Frequency (%) of use of common products and electrical appliances as reported in first questionnaire (1) and post miscarriage questionnaire (2)*

Products	Q	Daily/most days %	1/wk %	<1/wk %	Never %	Kappa	N
Disinfectant	(1)	33.0	31.5	20.2	15.3	0.29	202
	(2)	18 9	45.0	24.8	11.3		
Bleach	(1)	18.6	41.7	21.1	18.6	0.38	202
	(2)	13.1	39.4	33.0	14.5		
Window cleaner	(1)	2.0	8.9	48.3	40.9	0.54	202
	(2)	0.9	5.9	45.7	47.5		
Carpet cleaner	(1)	4.5	3.0	26.9	65.7	0.28	200
	(2)	2.7	4.5	24.3	68.5		
Oven/drain cleaner	(1)	3.0	5.4	30.0	61.6	0.36	202
	(2)	0.0	3.6	31.8	64.5		
Dry cleaning fluid	(1)	0.0	1 5	4.0	94.5	0.05	198
	(2)	0.0	0.0	3.6	96.4		
Turps/white spirit	(1)	0.0	2.0	16.8	81.2	0.40	201
	(2)	0.5	1.8	14 9	82.8		
Paint stripper	(1)	0.0	0 0	5.0	95.0	0.08	200
	(2)	0.0	0.0	4.1	95.9		
Paint/varnish	(1)	2.0	1.0	24.0	73.0	0.49	198
	(2)	1.4	1.8	17.7	79.1		
Weedkiller	(1)	0.0	0.0	6.5	93.5	0.29	200
	(2)	0.0	0.5	5.9	93.7		
Pesticides or insecticides	(1)	0.5	4.0	22.9	72.6	0.53	199
	(2)	0.9	3.2	19.1	76 8		
Aerosols or sprays	(1)	53.9	10.3	12.3	23.5	0.51	203
incl. hair spray	(2)	48.9	10.4	18.6	22.2		
Hair dyes/ bleach	(1)	1.0	0.5	13.0	85.5	0.57	199
	(2)	1.4	0.0	12.7	86.0		
Hair removal cream	(1)	0.0	2.0	10.0	88.0	0.69	198
	(2)	0.0	1.4	9 5	89.1		
Air freshener	(1)	33.2	17.3	17.3	32.2	0.44	200
	(2)	28.6	15.9	20.9	34.5		
Electric mixer/	(1)	4.9	9.8	37.1	48.3	0 43	203
liquidizer/grinder	(2)	3.2	10.4	34.4	52.0		
Vacuum cleaner	(1)	38.5	44.9	9.8	6.8	0.58	202
	(2)	38.0	41.2	15.4	5.4		
Floor polisher	(1)	0.5	2.4	1.5	95.6	0.54	203
	(2)	0.9	1.4	3.2	94.5		
Iron	(1)	26.7	50.5	18 0	4.9	0.59	203
	(2)	28.2	49.5	19.1	3.2		
Hair dryer/ other	(1)	50.7	24.4	9.8	15.1	0.61	201
elec. hair appliance	(2)	45.7	30.1	9.6	14.6		
Electric typewriter	(1)	3.9	5 4	10.2	80.5	0.66	203
	(2)	7.2	3.6	10.9	78.3		
Photocopier	(1)	22.5	5.4	7.8	64.2	0.82	203
	(2)	20.8	8.1	8.1	62.9		
PC or VDU	(1)	26.3	4.9	6.3	62.4	0.86	203
	(2)	26.2	9.5	4.5	59.7		
Power tool	(1)	0.0	0.5	6.3	93.2	0.45	204
	(2)	0.5	0.5	4.1	95.0		
Sunbed/sunlamp	(1)	0.0	2.4	1.0	96.6	0.76	203
	(2)	0.0	0.9	1.8	97.3		
Microwave oven	(1)	56.2	16.2	3.8	23.8	0.63	184
	(2)	53.6	16.4	5.0	25.0		

N is the number of women giving valid responses to the item in both questionnaires. Percentages are based on number of responses to each question in the individual questionnaires.

TABLE 2 *Frequency (%) of use of agents during work or leisure activity*

Agents	Q	Daily/most days %	1/wk %	<1/wk %	Never %	Kappa	N
Dental amalgam	(1)	1.0	0.0	0.5	98.5	0.80	199
	(2)	0.5	0.0	0.5	99.1		
Ceramic or enamels	(1)	0.5	0.5	0.5	98 5	0 39	204
	(2)	0.5	0.9	0.0	98 2		
Dry cleaning fluids	(1)	0 0	0.5	1.0	98.5	0.17	203
	(2)	0 9	0.9	1.8	96.4		
Glues	(1)	1.0	2.0	20.0	77.1	0 42	204
	(2)	0.5	2.3	7.7	89.6		
Leather work	(1)	0.0	0.0	0.5	99 5	1.00	204
	(2)	0.0	0.0	0.5	99.5		
Fabric or textiles	(1)	2.9	3.6	9.8	83.3	0.30	204
	(2)	3 6	3.6	4.5	88.3		
Dyes	(1)	0.0	0.0	2.4	97.6	0.18	205
	(2)	0.0	0.9	1.8	97.3		
Insecticides	(1)	0.0	2.4	9.8	87.8	0.31	204
	(2)	0 0	1 4	4.5	94.1		
Plastics	(1)	2.5	1.0	3.5	93.0	0.19	200
	(2)	0.9	0.9	0.5	97.7		
Metal cleaners or degreasers	(1)	0.5	4.9	12 3	82.3	0.20	203
	(2)	1.8	3.6	7.2	87.4		
Petrol	(1)	6.3	18.0	14.1	61.5	0.33	205
	(2)	1 8	6.3	13.5	78 4		
Paint	(1)	2.0	1.5	19.0	77.6	0.50	205
	(2)	1.8	2.3	12.2	83.8		
Photographic chemicals	(1)	0.5	1.5	1.0	97.1	0 45	205
	(2)	0.9	1.8	1.8	95.5		
Electrical wiring	(1)	1 0	1.0	3.9	94 1	0.34	204
	(2)	0.5	0.5	1.4	97.7		
Machining	(1)	1.5	1.0	1.5	96.1	0 45	203
	(2)	0.5	1 4	0 9	97.3		
Soldering	(1)	0 0	0.5	0.0	99 5	1.00	204
	(2)	0.0	0.5	0.0	99.5		
Radiation	(1)	2 0	0.0	2.5	95 6	0.63	203
	(2)	1.4	0.5	0.9	97.3		
Housework in other homes	(1)	5.9	4.9	4.9	84.3	0 84	204
	(2)	3.2	4.5	3 2	89.2		
Hairdressing	(1)	0.5	0.5	3.0	96.0	0.63	201
	(2)	0.5	0.0	2.3	97.3		
Farmwork	(1)	0 5	1.0	0.5	98.0	0.66	200
	(2)	0.5	0.5	0.0	99.1		
Hospital work	(1)	5.9	1.5	0.0	92.6	0.87	203
	(2)	5.9	2.3	0.5	91.4		
Shiftwork	(1)	7.0	2 2	0.0	90.8	0.76	185
	(2)	6.3	2.3	0.0	91.4		

N is the number of women giving valid responses to the item in both questionnaires. Percentages are based on number of responses to each question in the individual questionnaires.

use of common household products and electrical appliances from both questionnaires. Table 2 shows the overall percentages of women reporting frequency of use of particular work or hobby agents. Excellent agreement, defined as kappa ≥ 0.75, was found for the reporting of use of personal computer or visual display unit, photocopier, sunbed or sunlamp, dental amalgam, leather work, soldering, housework in other people's homes, hospital work and shiftwork. For those exposures with a low prevalence a high kappa value would be expected. Poor agreement, defined as kappa ≤ 0.40, was found for use of disinfectant, bleach, carpet cleaner, oven or drain cleaner, dry cleaning fluid, turpentine/white spirit, paint stripper, weedkiller, ceramic or

TABLE 3 *Percentage responses of the post miscarriage questionnaire compared with the first for frequency of use of household products and appliances*

Products/agents	Less frequent %	Same frequency %	More frequent %	*P*-value[a]
Disinfectant	32	45	23	0.10
Bleach	25	54	20	0.24
Window cleaner	17	73	9	0.02*
Carpet cleaner	19	66	16	0.40
Oven/drain cleaner	18	68	14	0.08
Dry cleaning fluid	6	91	4	0.20
Turps/white spirit	9	82	10	0.90
Paint stripper	4	93	4	0.82
Household paint	13	80	7	0.10
Weedkiller	5	92	4	0.84
Pesticides	11	81	8	0 30
Aerosols or sprays	28	56	16	0.02*
Hair dyes/bleach	6	89	5	0.89
Hair removal cream	4	93	3	0 65
Air freshener	28	52	21	0.08
Electric mixer	22	66	13	0.08
Vacuum cleaner	22	64	14	0.16
Floor polisher	1	96	3	0.44
Iron	14	70	16	0.52
Hair dryer or other	23	64	13	0.06
Electric typewriter	5	87	8	0.22
Photocopier	15	77	8	0.07
PC or VDU	8	83	9	0.64
Power tool	4	94	2	0.48
Sunbed/sunlamp	1	99	0	0.11
Microwave oven	24	66	10	0.02*

[a] 2-tail *P*-values are determined from the Wilcoxon signed rank test, *P*-values ≤ 0 05*.

enamels, fabric or textiles, dyes, insecticides, plastics, metal cleaners, petrol and electrical wiring. For the remaining variables the agreement was good with kappa values ranging from 0.42 (glue use) to 0.69 (hair removal cream).

Tables 3 and 4 present the data of the three new exposure frequency categories, 'less frequent', 'same frequency' or 'more frequent'. They show the comparison between the two questionnaires for household products and appliances and for agents at work and for leisure. The order of the variables in Tables 1–4 is the same as that found in the respective questionnaires.

In general the participants in this study reported less frequent use of household products, electrical appliances and agents during the first trimester of pregnancy in the post miscarriage questionnaire compared with the first one. Of the 48 questions, 34 were answered less frequently in the post miscarriage questionnaire. The binomial probability of finding 34 variables from 48 where the second answer is in the 'less frequent' category is 0.001.

When considering the differences in direction of answers between the questionnaires for individual variables, the reporting of less frequent use from data in the second questionnaire compared to information in the first questionnaire was statistically significant for window cleaner, aerosols or sprays, microwave oven, glues, insecticides, petrol, electrical wiring and work with radiation.

To try and explain the difference between the direction of the answers the data were further analysed by cross-tabulation with other key variables. There was no significant association with the woman's age and the post miscarriage response for any of the variables except use of air fresheners ($P = 0.0002$), oven or drain cleaner ($P = 0.05$) and use of household paint ($P = 0.05$). In the first two cases (Tables 5 and 6) the trend indicated that older women were more consistent with their replies to the two questionnaires. In the case of household paint (Table 7) older women showed an increase in replying 'less frequently' in the second questionnaire.

TABLE 4 *Responses of the post miscarriage questionnaire compared with the first for frequency of exposure to specific work or hobby products/agents*

Products/agents	Less frequent %	Same frequency %	More frequent %	*P*-value[a]
Dental amalgam	1	99	0	0.32
Ceramics/enamels	1	98	1	1.00
Dry cleaning fluids	1	96	3	0.07
Glues	15	83	2	< 0.001*
Leather work	0	100	0	1.00
Fabric or textiles	12	82	6	0.56
Dyes	2	96	2	0.58
Insecticides	9	88	2	0.02
Plastics	6	92	2	0.08
Metal cleaners	13	78	8	0.46
Petrol	28	66	5	< 0.001*
Paint	11	83	5	0.07
Photographic chemicals	1	96	2	0.36
Electrical wiring	4	95	1	0 04
Machining	3	96	1	0.16
Soldering	0	100	0	1.00
Radiation	3	97	0	0.02*
Housework in other homes	10	85	5	0.09
Hairdressing	2	98	1	0.23
Farmwork	1	99	0	0.18
Hospital work	0	98	2	0.28
Shiftwork	3	95	2	0.42

[a] 2-tail *P*-values are determined from the Wilcoxon signed rank test; *P*-values ≤ 0.05*.

TABLE 5 *Maternal age and repeat answer for frequency of use of air fresheners*

Air freshener use Age band	Less frequent n (%)	Same frequency n (%)	More frequent n (%)	Totals
<24	16 (49)	6 (18)	11 (33)	33
25–34	33 (27)	66 (53)	25 (20)	124
>35	5 (13)	28 (74)	5 (13)	38
Totals	54 (28)	100 (51)	41 (21)	195

$P = 0.0002$.

TABLE 6 *Maternal age and repeat answer for frequency of use of oven/drain cleaner*

Oven/drain cleaner use Age band	Less frequent n (%)	Same frequency n (%)	More frequent n (%)	Totals
<24	7 (21)	18 (55)	8 (24)	33
25–34	23 (18)	83 (66)	19 (15)	125
>35	5 (13)	33 (85)	1 (3)	39
Totals	35 (18)	134 (68)	28 (14)	197

$P = 0.05$.

TABLE 7 *Maternal age and repeat answer for frequency of use of household paint*

Household paint use Age band	Less frequent n (%)	Same frequency n (%)	More frequent n (%)	Totals
<24	4 (12)	29 (88)	0 (0)	33
25–34	13 (11)	97 (80)	12 (10)	122
>35	8 (21)	30 (79)	0 (0)	38
Totals	25 (13)	156 (81)	12 (6)	193

$P = 0.05$.

TABLE 8 *Number of previous miscarriages and repeat answer for frequency of use of electric iron use*

Electric iron use No. of previous miscarriages	Less frequent n (%)	Same frequency n (%)	More frequent n (%)	Totals
0	22 (17)	90 (70)	16 (13)	128
1–9	3 (5)	39 (70)	14 (25)	56
Totals	25 (14)	129 (70)	30 (16)	184

$P = 0.02$; Mantel Haenszel statistic for trend ($P = 0.005$).

There were no significant social class differences associated with repeatability of responses for the exposure variables. For those who had had one or more previous miscarriages compared to those who had no previous miscarriages, there was significantly more frequent use of an iron reported in the post miscarriage questionnaire ($P = 0.02$) and a significant Mantel-Haenszel trend test ($P = 0.005$) (Table 8).

The number of days between completion of the questionnaires ('time lag') varied between 7 days and more than 90 days. No statistically significant association was found with the direction of the answers from the post miscarriage questionnaire and the time lag between completion of the two questionnaires.

DISCUSSION

Although questionnaires were returned by 222 women, approximately 10% gave no information on exposures in their first questionnaire and so the analysis of repeatability is restricted to approximately 200 women. In general comparison between the responses to the two questionnaires showed a range of levels of agreement. For general home exposures (Table 1), 72% of the items had good to excellent agreement using the kappa statistic. For work or leisure exposures (Table 2) this figure was 57%. Excellent agreement (kappa >0.75) was more likely to be found for those variables where large numbers of women reported 'never' (low prevalence), for example for dental amalgam, soldering, leatherwork and sunbed. There were other situations where the kappa score was also excellent with significant numbers of women reporting frequent use, e.g. personal computer or visual display unit and photocopier. Where agreement was good (kappa >0.40, <0.75) there were several variables that had a high frequency of use (daily or most days) ranging between 30% and 50%. These included use of a microwave, hair dryer or other electrical hair appliance, iron, vacuum cleaner, aerosols or hair spray and air freshener. Low kappa scores (<0.40), representing the poorest agreement between the two responses, occurred where the distribution of responses was more evenly spread; examples included the use of bleach and disinfectant. In both cases their use may vary with month or season and household events. However, the lowest kappa scores were for frequency of use of dry cleaning fluids, dyes and plastics despite their being a substantial number of 'never used' responses. Although there is some dependency of kappa on prevalence of exposure the relationship is complex. There are examples of very low prevalence which give both excellent (use of sunbed) and poor (use of dry cleaning fluids) kappa results. These low and very low kappa scores may be explained by the nature of the multicategory ordinal questions, a result also found in a study examining the reproducibility of a questionnaire on risk

factors for osteoporosis.[9] It has been suggested that kappa coefficients may not be the most appropriate for assessment of agreement between categorical ordinal data where one pair of categories may be more dissimilar than another, and hence some instances of disagreement are worse than others. A pair of categories next to each other would have a greater agreement than a pair that was two categories apart but the kappa coefficient treats all instances of disagreement in the same way.[10]

The most important finding in this study was that when responses were discrepant, the post miscarriage response usually indicated a lower frequency of exposure than the first. This occurred for 34 of the 48 variables and in the case of both glues and petrol the category of 'less frequent' responses were significant at the 0.001 level. Analysis of the data shows that much of the change occurred from the 'less than once per week' category to 'never', i.e. from 'ever' to 'never'. This finding of less frequent reporting in the post miscarriage questionnaire may represent a bias which needs both explanation and further study. Although this is not a test-retest study, the two occasions for completing these questionnaires being separated by an adverse health outcome, the literature documenting such studies does indicate more 'no' responses on retest.[11,12] Similarly, assessment of the reproducibility of food frequency information as reported by 128 subjects over a time span of 6–12 months found the energy values determined from the second questionnaire to be lower than from the first for all 22 nutrients.[13] Van der Gulden *et al.*[14] studied occupational exposure of 209 men to eight specific compounds and reported the consistency of results between a questionnaire and a telephone interview conducted 3–5 weeks later. Of the changes in responses, 70% were from 'sometimes' to 'never', although different methods of collecting the information could have influenced the results.

The impact of second questionnaires reporting 'less frequent' exposures raises the issue of recall bias. Of 35 biases in analytical research,[15] five are 'possible sources of unequivalence in ascertaining or measuring exposure that appear directly related to the issue of recall'.[16] Mothers of children with birth defects are sometimes thought selectively to remember events during pregnancy and recall more than mothers of children without birth defects.[17] Klemetti and Saxen compared interview information before the sixth month of pregnancy with interviews after the pregnancy. Discrepancies between reports were not associated with pregnancy outcomes.[18] Mackenzie and Lippman also found no evidence to support differential reporting by women who had experienced adverse pregnancy outcomes.[19] Another study found no difference in reporting of exposures to drugs, smoking and environmental agents for women with adverse reproductive outcomes compared to those with a normal outcome. However, there was biased reporting of alcohol consumption: women with adverse reproductive outcomes reporting significantly less in the second interview compared with the first.[20] A study of autistic children also found no difference in accuracy of recall with respect to parental reporting.[21] Although recall bias could go in either direction: underreporting due to guilt or overreporting due to looking for a cause, the literature usually indicates that responses after an adverse outcome are interpreted as overreporting.

Recent studies indicate that recall bias may be due to controls (rather than cases) underreporting exposures.[22] In California, women who drank bottled water or abstained from drinking tapwater during early pregnancy had a lower than expected rate of spontaneous abortions and birth defects.[23] These studies followed a much publicized study of a leak of solvent into groundwater.[24] Subsequent work provided some confirmatory evidence[25] challenging the conventional view of recall bias. In other words it may not be those with the adverse outcome that are overreporting exposure but that those who have the normal outcome are underreporting.

We have considered possible explanations for the results presented here. These would include differences in perceived period of recall or a distortion of values or differences in rigour of recall, the latter two perhaps being directly influenced by the adverse outcome. Differences in the perceived period of recall could have occurred, although the exposure questions referred to a fixed period. The frequency of use of products or machines may be liable to change over time, either because of some obvious seasonality or because use may relate to specific events. Possibly women recalled the entire period when completing the first questionnaire but recalled only the latter part of the pregnancy when completing the post miscarriage questionnaire. However, it is not clear how this would result in systematic underreporting of responses in the second questionnaire.

Distortion of values for frequency of exposure or less rigorous recall might have been due to the answers in the post miscarriage questionnaire being influenced by the miscarriage. The experience of a miscarriage might have lead to a variation in respondents' motivation. It may have lead to either more rigorous searching for possible causes or to denial of any exposure to possible hazards. It may reflect some degree of 'wishful thinking', or feeling of guilt if the woman believed that the exposure was possibly harmful. The post

miscarriage questionnaire was sent with a letter of explanation and therefore women were aware of the study objective of finding causes of this adverse event. If this were the main source of change in the data between the two questionnaires, denial of exposure would have occurred. The miscarriage itself may also have resulted in a clinical depression with some loss of memory concerning the pregnancy including the period in question.

In this study the validity of frequency of use has not been tested because this would have required close surveillance of the individual to establish actual frequency of use. Without having a metered check on use of electrical appliances, regular home interview on use of household products, or a regular work visit to ascertain exposure to occupational agents it is not possible to validate these exposures. The validity of self-reported exposure has been considered by Stengel *et al.* using the assessment of experts as the gold standard. They found that for occupational exposure to solvents, sensitivity of the questionnaire was only 23–63% whereas the specificity was 87–98%.[26] In a study by Ahrens *et al.* information on asbestos exposure from a general questionnaire in a case-control study of lung cancer patients was compared with that from a supplementary and detailed questionnaire. They found that, in contrast to the work by Stengel *et al.*, the sensitivity for definite exposure was approximately 75%.[27] Orlowski *et al.* discuss the performance of supplementary detailed questionnaires which are job specific. In their view the general questionnaire may produce overestimates of exposure compared with a job-specific questionnaire.[28]

Clearly different responses will result in misclassification of exposure, which will in turn lead to different estimates of odds ratios. If the same investigation had been possible in a control group of women who had not had an adverse birth outcome, there may also have been a change in reporting towards less frequent use. Even if operating equally amongst women irrespective of outcome, memory loss would be important in interpreting retrospective studies. Such a memory loss would lead to a bias towards the null when calculating odds ratios.[29] Information bias investigated in a case-referent study of mental retardation and parental occupation found no evidence of differential misclassification. But the mean number of reported exposures and percentages of positive discordance suggested that referent parents rather than case parents, reported slightly more exposures.[30]

If recall bias really exists it might be expected that a second questionnaire would elicit an even greater number of responses towards less frequent use in the control group. Other investigators are encouraged to address this question by using a control group of subjects. In any case, this study suggests that women who have suffered an adverse pregnancy outcome do not appear to overreport their perceived exposure.

CONCLUSION

This study of the repeatability of responses of more than 200 women before and after a miscarriage found that use of chemicals and appliances was reported with varying consistency. Agreement on frequency of use of these items varied from excellent to poor. There was a tendency to report less frequent use after the miscarriage. This may be related to the particular adverse outcome and possible guilt feelings of the women or may be part of a general tendency to downplay past exposure or may be due to memory loss. Given the imperfect study design with no control group being sent a second questionnaire, the findings need to be put into perspective. It does however reinforce the need to be cautious when using the results from single surveys of retrospective self-reported exposure.

ACKNOWLEDGEMENTS

We are extremely grateful to all the mothers who took part and to the midwives for their co-operation and help in recruitment. The ALSPAC study team comprises interviewers, computer and laboratory technicians, clerical and research assistants, volunteers and managers who all continue to make the study possible. The project could not have been undertaken without the financial support of the Wellcome Trust, the Department of Health, the Department of the Environment, the Medical Research Council and the National Institute of Environmental Health Sciences (US).

REFERENCES

[1] Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; **70**: 213–20.

[2] Fleiss J L. *Statistical Methods for Rates and Proportion. 2nd Edn.* New York: John Wiley & Sons, 1977, pp. 212–36.

[3] Abramson J H. *Survey Methods in Community Medicine: An Introduction to Epidemiological and Evaluative Studies. 3rd Edn.* London: Churchill Livingstone, 1984.

[4] Fleiss J L. *Statistical Methods for Rates and Proportions.* New York: John Wiley & Sons, 1981.

[5] Landis J R, Koch G G. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.

[6] Brownson R C, Alavanja C R, Hock E T. Reliability of passive smoke exposure histories in a case-control study of lung cancer. *Int J Epidemiol* 1993; **22**: 804–08.

[7] Giovannucci E, Stampfer M J, Colditz G A *et al.* A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *Am J Epidemiol* 1993; **137:** 502–11.

[8] WHO. Recommended definitions, terminology and format for statistical tables related to the perinatal period. *Acta Obstet Gynecol Scand* 1977; **56:** 247–53.

[9] O'Neill T W, Cooper C, Cannata J B *et al.* Reproducibility of a questionnaire on risk factors for osteoporosis in a multicentre prevalence survey: The European Vertebral Osteoporosis Study. *Int J Epidemiol* 1994; **23:** 559–65.

[10] Maclure M, Willett W. Misinterpretation and misuse of the kappa statistic (Reviews and commentary). *Am J Epidemiol* 1987; **126:** 161–69.

[11] Collen M F, Cutler J L, Siegelaub A B, Cella R. Reliability of a self-administered medical questionnaire. *Arch Intern Med* 1969; **123:** 664–71.

[12] Mellner C. The self-administered medical history. *Acta Chir Scand* 1970; **Suppl.:** 406.

[13] Longnecker M P, Lissner L, Holden J M *et al.* The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. *Epidemiology* 1993; **4:** 356–65.

[14] Van der Gulden J W J, Jansen I W, Verbeek A L M, Kolk J J. Repeatability of self-reported data on occupational exposure to particular compounds. *Int J Epidemiol* 1993; **22:** 284–87.

[15] Sackett D L. Bias in analytical research. *J Chron Dis* 1979; **32:** 51–63.

[16] Lippman A, Mackenzie S G. What is recall bias and does it exist? In: Marois M (ed.). *Prevention of Physical and Mental Congenital Defects. Part C: Basic and Medical Science, Education and Future Strategies*. New York: Alan R Liss Inc, 1985.

[17] Jick H. Recall 'error' in interview studies of past drug use. *Am J Public Health* 1979; **72:** 405.

[18] Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformation. *Am J Public Health* 1967; **57:** 2071–75.

[19] Mackenzie S G, Lippman A. An investigation of report bias in a case control study of pregnancy outcome. *Am J Epidemiol* 1989; **129:** 65–75.

[20] Feldman Y, Koren G, Mattice D *et al.* Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. *Teratology* 1989; **40:** 37–45.

[21] Deykin E Y, MacMahon B. Pregnancy, delivery and neonatal complications among autistic children. *Am J Dis Child* 1980; **134:** 860–64.

[22] Neutra R R, Swan S H, Hertz-Piciotto I *et al.* Potential sources of bias and confounding in environmental epidemiologic studies of pregnancy outcome. *Epidemiology* 1992; **3:** 134–42.

[23] Wrensch M, Swan S H, Lipscomb J, Epstein D M, Neutra R R, Fenster L. Spontaneous abortion and birth defects related to tap and bottled water use, San Jose Calif 1980–85. *Epidemiology* 1992; **3:** 98–103.

[24] Deane M, Swan S H, Harris J A, Epstein D M, Neutra R R. Adverse pregnancy outcomes in relation to water consumption: a reanalysis of data from the original Santa Clara County Study, California 1980–1981. *Epidemiology* 1992; **3:** 94–97.

[25] Fenster L, Windham G C, Swan S H, Epstein D M, Neutra R. Tap or bottled water consumption in a case-control study of reporting consistency. *Epidemiology* 1992; **3:** 120–24.

[26] Stengel B, Pisani P, Limasset J, Bouyer J, Berrino F, Hemon D. Retrospective evaluation of occupational exposure to organic solvents: Questionnaire and job exposure matrix. *Int J Epidemiol* 1993; **22:** S72-S82.

[27] Ahrens W, Jockel K, Brochard P *et al.* Retrospective assessment of asbestos exposure. I. Case-control analysis in a study of lung cancer. *Int J Epidemiol* 1993; **22:** S83-S95.

[28] Orlowski E, Pohlabeln H, Berrino F *et al.* Retrospective assessment of asbestos exposure. II. At the job level: Complementarity of job-specific questionnaire and job exposure matrices. *Int J Epidemiol* 1993; **22:** S96-S105.

[29] Copeland K T, Checkoway H, McMichael A J *et al.* Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105:** 488–95.

[30] Roeleveld N, Kiemeney L, Schattenberg G, Peer P. Information bias in a case-referent study on mental retardation and parental occupation: colleagues as dual respondents. *Epidemiology* 1990; **1:** 293–97.

(Revised version received November 1995)