AMERICAN JOURNAL OF EPIDEMIOLOGY
Vol. 129, No. 1
Printed in U.S.A.

AN INVESTIGATION OF REPORT BIAS IN A CASE-CONTROL STUDY OF PREGNANCY OUTCOME

SUSAN G. MACKENZIE[1] AND ABBY LIPPMAN[2]

Mackenzie, S. G., and A. Lippman (Dept. of Epidemiology and Biostatistics, McGill U., Montreal, Quebec H3A 1A2, Canada). An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989;129:65–75.

The role of report (recall) bias in case-control studies of possible reproductive hazards was investigated in a study of women who gave birth at the Royal Victoria Hospital, Montreal from September 1983 to May 1985. Women were questioned twice (early in pregnancy; after delivery) about exposures that might influence pregnancy outcome. The two sets of responses of case mothers, control mothers, and mothers of infants of intermediate health status were then compared. Similar inconsistencies in the reporting of 39 exposure variables were common in all three groups, with postdelivery deletion of previous reports more frequent than addition of new information. Changes in reporting were not associated with pregnancy outcome, maternal concern about the baby or maternal sociodemographic characteristics. Odds ratios of exposure estimated from the two sets of data did not differ importantly. Moreover, there was no postdelivery trend to increases, or decreases, in the estimates of the odds ratios. The data do not provide evidence of biased reporting of exposures.

epidemiologic methods; pregnancy; retrospective studies

In case-control studies of reproductive outcomes, recall bias is often mentioned as a potential cause of exposure misclassification. This bias is said to result from cases searching more carefully for possible causes of their infants' illnesses and being, therefore, more likely to recall (and report) exposures than are equally exposed healthy controls.

Given the number and variety of influences on recall and reporting, and their potential to interact with each other (1–6), it is not difficult to construct scenarios in which cases and controls *might* be differentially influenced by events and might produce reports of differing validity. Consequently, although little is actually known about biased recall (or reporting), expressions of concern about its possible effects are not difficult to find in the literature

Received for publication September 29, 1987, and in final form April 14, 1988.

[1] Department of Community Health, Lakeshore General Hospital, Pointe Claire, Quebec, Canada.

[2] Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Ave. West, Montreal, Quebec, H3A 1A2, Canada. (Send reprint requests to Dr. Abby Lippman at this address.)

The research was funded by the National Health Research Development Program of Health and Welfare Canada. Dr. Mackenzie was supported by fellowships from the Fonds de la recherche en santé du Québec and from the National Health Research Development Program. Dr. Lippman is supported by the National Health Research Development Program through a National Health Research Scholar Award.

This work was done in partial fulfillment of the requirements for the degree of Doctor of Philosophy awarded to Dr. Mackenzie.

The authors thank Drs. Jean-François Boivin, James Hanley, and Antonio Ciampi (Department of Epidemiology and Biostatistics, McGill University) and Dr. Robert Usher and Ms. Frances McLean (Neonatal Unit, Royal Victoria Hospital) for helpful discussions; Dr. Paige Kaplan (Montreal Children's Hospital) for diagnostic assistance; and Ms. Micheline Wheeler, Ms. Mary McLenahan and Ms. Susan Munday for their help in data collection.

(see, for example, references 7–10). Yet, despite calls for investigations of the nature and importance of the phenomenon (11–14), little experimental evidence exists to support or refute any of the warnings and concerns about recall bias.

Most of the literature about recall bias derives from research designed for other purposes in which the "reliability" or "validity" of participants' reports was apparently investigated as an afterthought. None of these indirect studies of the existence and importance of report bias is entirely satisfactory, mainly because information obtained directly from participants in case-control studies was compared with data from other sources—often health records or the reports of proxy respondents—and problems other than recall bias could explain any discrepancies found. Only one study that specifically addresses the subject has been cited with any frequency. And, unfortunately, this work by Klemetti and Saxén (15) has fairly often (six of 31 citations listed in the *Science Citation Index* by April 1986) been incorrectly said to provide evidence of the existence of report bias despite the authors' clearly stated conclusion that [in explaining] "... the great discrepancy between prospective and retrospective 'memory' ... the outcome of the pregnancy and the condition of the child born do not play a major role" (15, p. 2074).

Given the limited data available to evaluate rigorously the nature and extent of recall bias, it seemed useful to study it directly again. The present "case-control within a cohort" (or "nested case-control") study (16) was designed specifically to examine this potential bias. Women identified in early pregnancy were asked about their recent exposures and their reproductive, family, and medical histories on a specially developed self-administered early pregnancy questionnaire. All women who subsequently gave birth to infants who were stillborn, died, or had serious health problems (cases), all women whose infants had less serious problems (called in this study "neither-case-nor-controls"), and a random sample of women who gave birth to healthy infants (controls) were sent a second questionnaire soon after delivery. This postdelivery questionnaire included exactly the same questions about exposures and experiences during pregnancy as the early pregnancy questionnaire. To reveal the effect of knowledge of the outcome of pregnancy on the reporting of exposures and experiences, early pregnancy and postdelivery exposure reports were compared, as were the corresponding measures of association of the exposures and pregnancy outcome. If report bias operates as it is said to, cases would be expected to add more new reports of exposure, and to delete less of the previously reported information, on the second occasion, and estimates of the odds ratios should reflect these differential changes.

It should be noted that, in this work, we refer to "report" rather than to "recall" bias. Although there may be differences in the completeness of data provided by cases and controls in a given study, only a respondent's *reports* are available to us, and we can rarely, if ever (under ordinary survey conditions), determine whether something has influenced the recall of exposure or the reporting of the recalled information. Since the more general term "report bias" provides a better description of the underlying phenomena, it is the one we have chosen to use.

MATERIALS AND METHODS

The target population for the study comprised women who intended, from early pregnancy, to give birth at a large urban hospital (the Royal Victoria Hospital, Montreal) between September 1, 1983 and April 30, 1985. Women in the target population who were literate in either French or English and had had their first visits for antenatal care before the 16th week of pregnancy were invited to participate in the study, either at their obstetricians' offices or at the antenatal blood test clinic of the Royal Victoria Hospital. Each woman who expressed interest in the study was given

an enrollment package that contained a letter that explained the study, the (self-administered) early pregnancy questionnaire, and a postage paid return envelope.

The early pregnancy questionnaire contained questions on 39 exposure variables that are either common exposures (such as smoking, contraceptive use, and medications) or are generally relevant to studies of reproductive outcomes. A complete list of the exposures studied may be found in table 2. To help prevent errors in reporting of exposure due to erroneous recall of the *time* of exposure, inquiries about most exposures were directed to the time between a woman's last menstrual period and her first visit to her doctor for antenatal care—dates that are important to pregnant women. Questions about contraception referred to the six months before the last period and those about consumption of coffee, alcohol, and cigarettes referred to the month before the last period. All reports were scored dichotomously (exposed or unexposed). Copies of the questionnaire are available on request.

Women who did not return a completed early pregnancy questionnaire within two weeks were telephoned and urged to respond. Only those who responded before the start of the 20th week of pregnancy were eligible to continue in the study.

Participants who gave birth at the Royal Victoria Hospital were identified from hospital records, and the charts of their infants were reviewed. Women who gave birth to twins, those whose deliveries were ascertained more than two weeks after the infant's birth, those for whom the dating of the pregnancy was inconsistent with information from the early pregnancy questionnaire, and those who gave birth to infants whose health could not be assessed with confidence from information in their charts were ineligible for the postdelivery part of the study and were not followed further.

Women who were eligible for the postdelivery part of the study were classified into three groups on the basis of the health and hospital accommodation (neonatal intensive care unit vs. normal nursery) of their infants.

To be a case, a woman had to have given birth to an infant who met *at least one* of the following criteria: 1) the infant was stillborn or died before discharge from the hospital; 2) the infant was admitted to the neonatal intensive care unit for more than 24 hours or was transferred for care to the Montreal Children's Hospital; 3) the infant was born with a major malformation (that was diagnosed or suspected before discharge from hospital), or had another serious problem, even if intensive care was not required in the neonatal period.

To be a control, a woman had to have given birth to a liveborn infant who met *all* of the following criteria: 1) the infant was neither admitted to the neonatal intensive care unit nor transferred to the Montreal Children's Hospital; 2) the gestational age was between 37 and 42 weeks, and the birth weight was between the 3rd and 97th percentiles for gestational age; 3) the infant had no significant health problems while in hospital.

A woman was classified into the intermediate neither-case-nor-control group if her infant's condition made her ineligible to be a control but the problem was not serious enough for her to be a case. (Although such a group would not usually be defined in a case-control study, its inclusion here would, if there were any effect of pregnancy outcome on reporting of exposures, permit the data to be examined for evidence of a "dose-response" relation.)

All women in the case group and the neither-case-nor-control group were sent a postdelivery questionnaire two weeks after they had given birth, as were 60 per cent (randomly selected) of the women in the control group. The postdelivery questionnaire repeated the questions about early pregnancy experiences and exposures and reproductive and family histories from the early pregnancy questionnaire. It also included questions about the course of pregnancy subsequent to completion of the early pregnancy questionnaire, about labor

and delivery, and about the infant. To assess maternal concern about the infants' problems, each woman was also asked if her baby had any malformation or other problem and, if so, whether she felt "not concerned, somewhat concerned, moderately concerned, or very concerned" about it.

The consistency of the early pregnancy and postdelivery reports was assessed by calculating kappa statistics for the pairs of reports for each exposure (17). (Women who failed to answer a given question on one or both questionnaires were excluded from the analysis of the responses to that question.) To determine if inconsistency was due to reports of more exposure or of less exposure on the postdelivery questionnaire, specific changes were examined, with the early pregnancy questionnaire reports arbitrarily chosen as the standards. Four combinations of reports were possible: 1) they could be consistently positive (exposed); 2) they could be consistently negative (unexposed); 3) they could be inconsistent when exposure was reported on the early pregnancy questionnaire but not the postdelivery questionnaire—a deletion; or 4) they could be inconsistent when exposure was reported on the postdelivery questionnaire but not the early pregnancy questionnaire—an addition. The frequencies of additions and deletions were determined for each of the 39 exposures studied. The frequency of addition (deletion) was the number of additions (deletions) divided by the number of early pregnancy questionnaire-unexposed (early pregnancy questionnaire-exposed) reports.

The effects of a number of potential determinants of changes in reporting were directly assessed by examining their associations with the frequencies of addition and deletion of exposure reports. Factors studied in this way were: pregnancy outcome; maternal education; country of birth; occupation; outcome of the preceding pregnancy; self-perceived risk of having a study baby with a problem (early pregnancy report); maternal concern about the study baby (postdelivery report); and, finally, the number of days between a woman's response to the two questionnaires. This part of the analysis was restricted to exposure variables for which there had been at least 25 additions ($n = 13$ exposures) or deletions ($n = 10$ exposures).

Although individual changes were interesting, our major aim was to investigate the effect of these changes on measurements of exposure and on measures of association. Only if the conclusions about exposure/disease associations based on postdelivery questionnaire responses differed from those based on the early pregnancy questionnaire would concerns about biased reporting be warranted.

To study the consistency of the estimates of exposure prevalences and measures of association, all responses provided were used—whether or not questions were answered on both questionnaires. Exposure prevalences were estimated separately from early pregnancy and postdelivery data for each of the three pregnancy outcome groups. The statistical significance of the associations of questionnaire (early pregnancy vs. postdelivery) and exposure prevalence, and the homogeneity of the associations across the three pregnancy outcome groups, were assessed using chi-square tests (17). The associations of the exposure variables and pregnancy outcome were also assessed from both early pregnancy and postdelivery data. Odds ratios for case/control and neither-case-nor-control/control comparisons were estimated from standard 2 × 2 tables, and their approximate 95 and 99 per cent confidence intervals were calculated with the use of Miettinen's test-based method (18).

RESULTS

The study population

At the hospital, the early pregnancy questionnaire was offered to 1,352 eligible women; 134 (9.9 per cent) refused to participate. Of the 1,218 women who accepted the questionnaire, 756 (62.1 per cent) subsequently returned completed early pregnancy questionnaires. Of the question-

naires distributed at obstetricians' offices, 661 were returned completed. Unfortunately, because distribution of these questionnaires was handled by office (rather than study) personnel, neither the frequencies of ineligibility or refusal to participate, nor the total number of early pregnancy questionnaires distributed was known for most practices. The response rate for questionnaires distributed at offices that were able to keep accurate records was 60.9 per cent.

Among the 1,417 early pregnancy questionnaire respondents, 131 (9.2 per cent) were not eligible to participate in the study (primarily because they answered the early pregnancy questionnaire after 20 weeks (n = 114) and/or had had their first antenatal visit after 16 weeks or refused consent for follow-up (n = 12)). A further 124 early pregnancy questionnaire respondents were subsequently excluded from the study because they did not give birth at the Royal Victoria Hospital, with more than a third of this latter group (n = 47) having had spontaneous abortions.

Of the 1,162 eligible early pregnancy questionnaire respondents who gave birth at the hospital as planned, 76 failed to meet the eligibility criteria for the postdelivery part of the study (11 women gave birth to twins; 27 had infants whose health could not be assessed with confidence from information in the infants' charts; for 23, the dating of the pregnancy in the infant's chart was not consistent with that reported on the early pregnancy questionnaire; and 15 deliveries were ascertained more than two weeks after the baby's birth).

Representativity of study participants

Early pregnancy questionnaire respondents eligible for the study were compared with the 3,465 women in the target population who gave birth between April 1, 1984 and March 31, 1985. The women in the two groups were similar with respect to age, previous reproductive history, and outcome of the current pregnancy, but the study women were more likely to have been born in Canada and to be highly educated. Their current infants weighed more, were less often born prematurely and less frequently required intensive care than those of women in the target population. These differences could be explained by 1) exclusion from the study of women who were not literate in either English or French (mainly women from the large immigrant population served by the Royal Victoria Hospital) and 2) selective refusal of less educated women to participate in the study if they found it too onerous to read and respond to a 12-page questionnaire. Whatever the reasons for it, the higher socioeconomic status of study women may underlie their lower frequencies of premature and low birth weight infants.

Pregnancy outcome and response to the postdelivery questionnaire

Of the 1,086 women who were eligible for the postdelivery part of the study, 71 per cent were classified as controls, 8 per cent as cases, and 21 per cent as neither-case-nor-controls. The overall response rate to the postdelivery questionnaire, which did not vary between the three groups, was 95.5 per cent, and the main study group comprised 747 women (445 controls, 217 neither-case-nor-controls, and 85 cases). The mean intervals between delivery and our receipt of the completed postdelivery questionnaire were 33.5 (±14.9), 33.1 (±14.1) and 35.2 (±16.0) days for women in the control, neither-case-nor-control, and case groups, respectively.

Problems of the infants

The health problems experienced by infants in the case and neither-case-nor-control groups are summarized in table 1. Intensive care was required for 87.1 per cent of cases and 39.2 per cent of neither-case-nor-controls. (The 11 cases not admitted to the neonatal intensive care unit included five who were stillborn, two who were transferred directly to the Montreal Children's Hospital and four whose problems, though important, did not require

TABLE 1

Health problems of the case and neither-case-nor-control infants: report bias study, Montreal, Canada, 1983–1985

Problem	Cases (n = 85)		Neither-case-nor-controls (n = 217)	
	No.	(%)	No.	(%)
Stillborn	5	(5.9)		
Liveborn				
With malformations, by gestational age (weeks)				
<37	3	(3.5)		
≥37	10	(11.8)	41	(18.9)
Without malformations by gestational age (weeks)				
<37				
No problem(s)	15	(17.6)	10	(4.6)
Problem(s)	4	(4.7)	4	(1.8)
>42				
No problem(s)			11	(5.1)
Term, small for gestational age*	3	(3.5)	2	(0.9)
Term, large for gestational age†			25	(11.5)
Phototherapy only	1	(1.2)	24	(11.1)
In neonatal intensive care unit for observation‡	14	(16.5)	46	(21.2)
Other	30	(35.3)	54	(24.9)

* Small for gestational age, birthweight below the 3rd percentile for gestational age.

† Large for gestational age, birthweight above the 97th percentile for gestational age.

‡ Most of these infants were admitted to the neonatal intensive care unit because of (possible) aspiration of amniotic fluid or meconium, because their mothers had elevated temperatures during labor, or because of (relatively) minor respiratory problems.

intensive care.) The median lengths of stay in the neonatal intensive care unit were four days for the cases and nine hours for the neither-case-nor-controls, which reflects the generally greater severity of the problems of the cases.

Although fewer case than neither-case-nor-control infants were malformed (15 per cent vs. 19 per cent), the malformations of the cases were generally more serious. The 13 case infants with malformations included three with congenital heart defects, three with dislocated hip(s) (including one who also had auricular pits), and one each with a café au lait spot, a hydrocele and hernia, a meningomyelocele, positional skeletal abnormalities, Potter syndrome, a renal malformation, and trisomy 18. The 41 neither-case-nor-control infants with malformations included eight with positional abnormalities of the lower limbs, seven with dislocatable hips, seven with birth marks or moles, five with hypospadias, three with extra fingers, three with hydroceles, three with toe abnormalities, two with umbilical hernias and one each with a cleft gum, an ear tag, and intrauterine pressure deformities.

As expected, maternal concern increased from the control through the neither-case-nor-control to the case group, and the association of concern with pregnancy outcome was highly statistically significant ($p < 0.001$).

Overall agreement of and specific changes in the reports of individual women

Although the degree of agreement between early pregnancy and postdelivery reports was always better than would be expected on the basis of chance alone, there was, nevertheless, considerable inconsistency of reporting. The frequencies of addition and deletion for each of the 39 exposure variables and the values of kappa are shown in table 2. When the kappa values were grouped as suggested by Landis and Koch (19) to describe the degree of agreement, there was "fair agreement" (kappa = 0.21–0.40) for five exposures, "moderate agreement" (kappa = 0.41–0.60) for 13, "substantial agreement" (kappa = 0.61–0.80) for 14, and "almost perfect agreement" (kappa = 0.81–1.00) for the remaining seven exposures. Values that indicated only "slight agreement" (kappa = 0–0.20) or "poor agreement" (kappa <0) were not observed. As one might expect, reports about previous pregnancies were more consistent than those about transient experiences that occurred early in the study pregnancy.

For most of the exposure variables, the

TABLE 2

Changes in reports of exposure provided on the early pregnancy and postdelivery questionnaires by women in the three pregnancy outcome groups: control, neither-case-nor-control, and case: report bias study, Montreal, Canada, 1983–1985

Exposure	Consistency of postdelivery reports of exposure three outcome groups combined			Prevalence (%) calculated from questionnaires (early pregnancy/postdelivery), by outcome group			Chi-square
	Addition (%)	Deletion (%)	Kappa	Control (n = 445)	Neither-case-nor-control (n = 217)	Case (n = 85)	
Nausea	8.1	20.9	0.63	72.3/59.3	73.5/59.6	61.5/45.6	***
Poor nutrition	13.2	38.8	0.49	46.4/36.0	45.7/31.9	33.8/34.1	***
Coffee	13.1	8.3	0.78	68.9/67.0	60.6/60.9	66.7/69.6	
Wine	22.7	4.5	0.75	75.6/80.0	71.0/72.8	78.6/75.0	
Liquor	14.9	14.0	0.70	43.5/45.3	37.0/41.6	37.8/42.5	
Smoking	1.3	4.4	0.95	36.5/35.9	30.7/30.1	33.3/34.1	
Coffee decrease	8.8	27.4	0.64	50.8/42.5	49.3/38.7	42.5/32.9	***
Wine decrease	14.4	26.0	0.59	53.1/50.1	54.7/41.1	55.7/41.0	**
Liquor decrease	11.7	23.2	0.65	35.0/34.5	32.5/32.1	27.3/27.3	
Smoking decrease	3.6	22.6	0.77	33.4/23.3	25.6/22.1	24.3/21.0	*
Radiation	2.6	26.0	0.68	8.6/8.4	5.7/7.8	2.4/3.8	
Other fetal danger	6.3	50.9	0.46	15.7/13.8	14.3/11.3	18.1/10.1	
Emotional lability	1.2	77.1	0.30	7.8/3.3	6.1/1.9	4.8/1.2	***
Other stress	5.7	40.4	0.58	22.9/18.4	24.6/19.7	16.9/12.5	*
Chronic illness	4.6	45.9	0.55	20.2/13.3	21.8/16.7	20.5/18.8	**
Cold	2.6	67.4	0.38	12.8/7.4	13.5/4.5	11.9/6.2	***
Ear, nose, or throat/respiratory illness	0.9	61.0	0.49	5.7/3.0	5.6/3.0	4.8/3.7	*
Headache	0.3	65.5	0.48	4.1/1.4	3.7/2.5	4.8/1.2	*
Gastrointestinal illness	1.9	76.3	0.27	5.3/3.2	4.2/1.5	8.3/6.2	*
Other acute illness	3.4	55.2	0.27	12.6/7.6	14.4/10.9	15.5/12.3	*
Aspirin or acetaminophen	4.4	46.9	0.54	18.4/13.7	16.7/12.9	22.0/9.9	**
Vitamins	1.5	73.7	0.32	6.0/3.5	2.8/1.9	7.3/2.5	**
Cold medication	1.2	52.6	0.56	8.0/5.6	7.4/2.4	8.5/7.4	*
Antibiotics or antibacterials	0.7	34.1	0.73	5.7/4.2	5.1/4.3	6.1/6.2	
Other medication	3.0	43.8	0.60	20.0/13.7	18.1/12.6	19.5/11.1	**
Non-medical drugs	0.4	23.1	0.81	3.4/3.0	4.7/3.9	2.5/2.5	
Oral contraceptives	2.1	26.6	0.75	12.8/11.2	11.7/11.6	20.0/15.0	
Vaginal spermicides	1.4	15.8	0.83	8.7/9.0	8.0/6.3	3.7/6.2	
Previous liveborn child	1.6	1.7	0.97	48.3/49.5	49.3/49.3	36.9/37.4	
Previous spontaneous abortion	1.0	4.8	0.94	14.7/15.2	13.0/11.8	16.7/19.3	
Previous induced abortion	2.2	11.5	0.84	9.3/10.1	13.0/14.2	11.9/10.8	
Previous other pregnancy outcome	0.4	19.0	0.82	2.7/2.3	3.3/3.8	3.6/2.4	
Previous malformed child	0.3	35.7	0.72	2.0/1.6	1.4/1.4	2.4/1.2	
Previous premature child	1.0	47.6	0.55	1.8/1.8	4.6/3.3	3.6/3.6	
Previous child with other problem	1.3	38.7	0.63	5.0/4.1	3.7/3.8	1.2/2.4	
Maternal sibling with malformation	0.6	26.9	0.77	3.7/3.1	3.8/3.9	2.4/2.6	
Maternal sibling with other problem	2.5	52.0	0.42	3.4/4.7	3.3/2.9	3.7/3.8	
Paternal sibling with malformation	0.9	25.0	0.66	0.7/0.9	2.4/3.5	5.0/5.1	
Paternal sibling with other problem	1.8	39.4	0.60	4.2/5.4	4.9/3.5	6.2/2.6	

Statistical significance of the overall chi-square test of the differences of the prevalences reported on the two questionnaires: * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$.

frequency of deletion was considerably higher than that of addition. Net changes (number of additions − number of deletions) were negative for 29 variables and positive for seven; three showed no change. The high frequency of deletion was particularly evident for reports of illnesses and of medication use. Additions were most fre-

quent for nausea, poor nutrition, consumption of coffee, wine, and liquor in the month before the last menstrual period, and for subsequent decreased consumption of the same substances.

The effects of potential determinants of changes in reporting on the frequencies of addition and deletion

When the associations of selected maternal and other factors (see Materials and Methods) and changes in reporting were studied, 80 tests for associations with deletions and 96 for associations with additions were done. Each set of tests yielded six results that were statistically significant at the 0.05 level. However, only three of the factors were associated with changes in reporting of more than one exposure—and these effects were not consistent. Given the rarity of statistically significant associations and the nonuniformity of the observed effects, it does not appear that any of these factors can account for either deletion or addition.

Changes in exposure levels calculated for the three pregnancy outcome groups

As expected from the excess of deletions over additions, postdelivery prevalences tended to be lower than those calculated from early pregnancy in all three pregnancy outcome groups (see table 2). Postdelivery decreases in prevalence were statistically significant for 17 exposure variables—mostly transient experiences of early pregnancy; no statistically significant increases in prevalence were observed. Chi-square tests of heterogeneity for the associations of prevalence and time of questioning (early pregnancy versus postdelivery) revealed no statistically significant differences among the three pregnancy outcome groups for any of the 39 exposure variables.

Consistency of estimates of the measures of association

Although the frequency of prevalence changes did not vary significantly according to pregnancy outcome, inconsistencies in the estimates of the measures of association from the early pregnancy and postdelivery data in the three pregnancy outcome groups might still have provided evidence of biased reporting if the postdelivery estimates of the odds ratios were consistently higher (or lower). We therefore compared the early pregnancy and postdelivery odds ratios from case/control and neither-case-nor-control/control comparisons for each of the 39 exposures. The postdelivery estimates of the odds ratios were not generally higher (or lower) than their early pregnancy counterparts: postdelivery increases were seen for only 17 of the 39 case/control comparisons and 19 of the 39 neither-case-nor-control/control comparisons.

More importantly, there was a statistically significant association of exposure and pregnancy outcome (the 95 per cent confidence interval did not include 1.0) on one but not the other questionnaire for only three of the 39 case/control comparisons of early pregnancy and postdelivery results (nausea, poor nutrition, and previous liveborn child) and only five of 39 neither-case-nor-control/control comparisons (coffee, wine, wine decrease, previous premature child, and malformation of a paternal sibling) (see table 3). As can be seen, however, the differences in the case/control findings for previous liveborn child are of no real importance, and the other differences are hardly more impressive. Moreover, had we chosen to use 99 per cent confidence intervals in view of the multiple comparisons, then the conclusions about the associations of the exposures and pregnancy outcome would have been the same for all 39 exposures studied. The only exposure significantly associated with pregnancy outcome on both questionnaires was malformation of a paternal sibling in the case/control comparisons; odds ratios (and 99 per cent confidence intervals) from the early pregnancy and postdelivery questionnaires were 7.5 (1.4–41.5) and 5.7 (1.1–29.6), respectively.

TABLE 3

Exposures for which estimates of the odds ratios (ORs) calculated from early pregnancy and postdelivery data differed: report bias study, Montreal, Canada, 1983–1985

Groups compared	Exposure	Early pregnancy		Postdelivery	
		OR	(95% CI*)	OR	(95% CI)
Case/control	Nausea	0.61	(0.37–1.01)	0.57	(0.35–0.93)
	Poor nutrition	0.59	(0.36–0.97)	0.92	(0.56–1.50)
	Previous liveborn child	0.63	(0.39–1.01)	0.61	(0.38–0.98)
Neither-case-nor-control/control	Coffee	0.69	(0.49–0.98)	0.77	(0.54–1.08)
	Wine	0.79	(0.54–1.14)	0.66	(0.46–0.99)
	Wine decrease	1.07	(0.76–1.50)	0.69	(0.49–0.98)
	Previous premature child	2.66	(1.07–6.62)	1.81	(0.66–5.00)
	Paternal sibling malformed	3.56	(0.92–13.77)	3.77	(1.18–12.02)

* CI, confidence interval.

DISCUSSION

Our observations that many exposures, particularly the transient exposures of early pregnancy, were reported less frequently on a second postdelivery questionnaire, but that the changes in reporting observed were not associated with pregnancy outcome, agree with those from the one previously published study designed to directly test the consistency of early pregnancy and postdelivery reporting (Klemetti and Saxén (15)). Three important design features that distinguish the present study from much previous work strengthen the conclusion that report bias may not be as serious a problem as intuition would suggest. 1) Obtaining both sets of exposure reports from the study participants, also a feature of the study of Klemetti and Saxén, eliminated potential origins of discrepant information inherent in comparisons of self-reports with data from other sources. 2) The use of self-administered questionnaires, unique to this study, ensured the exact replication of all questions to all women on both occasions and precluded the possibility of interviewer bias in data collection. Since this approach also left women free to use whatever resources they felt appropriate in answering the questions, it increased both the opportunity for case mothers to report as much exposure information as possible on the postdelivery questionnaire, and our ability to detect any resultant reporting bias. 3) Mailing the postdelivery questionnaires shortly after women had given birth introduced a measure of control over the passage of time not present in earlier studies in which varying, and sometimes long, periods of time elapsed between delivery and collection of exposure reports. For instance, more than half of the postdelivery interviews in the study of Klemetti and Saxén were done by the end of the third postdelivery month but the rest were not completed until 15 months after the women had given birth.

The large number of exposures we examined, 39 variables versus only two studied by Klemetti and Saxén, also strengthen our study and permitted us to explore the extent to which conclusions about changes in reporting could be generalized. By evaluating postdelivery questionnaire changes in the reports of *all* women, and studying major changes (addition and deletion of reports), we were able to assess the net effects of the changes in reporting on both the exposure prevalences and the measures of association. In contrast, Klemetti and Saxén appear to have studied only the frequency of "nonidentical" postdelivery reports in the subset of women who had reported exposure at the early pregnancy interviews.

Our failure to demonstrate report bias is especially notable, since inescapable fea-

tures of the design—as well as events beyond our control—would, if anything, have increased the opportunity for report bias. For example, because of their infants' problems, some case and neither-case-nor-control mothers might have been questioned about exposures by health care providers before the postdelivery questionnaire was answered. Or, the postdelivery questionnaire might have been answered with less care (and less accuracy, leading to less consistency) than the early pregnancy questionnaire, either because the women were generally less interested in the questions when the pregnancy was over or because they recognized that they had already given us the information and were not motivated to answer a second time. These situations provided an opportunity for differential reporting according to pregnancy outcome, yet, the data do not provide evidence that it occurred. In a slightly different vein, the use of self-administered questionnaires, with the consequent absence of interviewers, precluded the use of probing techniques that can increase both the quality and quantity of information reported (2, 3). In addition, the rather liberal use of open questions may have led to reports being generally incomplete (20). These last two features could have led to underreporting of exposures, even on the early pregnancy questionnaire, thus further enhancing the opportunity for case and neither-case-nor-control mothers to report more new exposures on the postdelivery questionnaire than the controls.

Since we did not observe report bias, the possibility of a type II error must be considered. Because fewer study participants than anticipated were classified as cases, the confidence intervals for the estimates of the odds ratios were wide, making it difficult to conclude that the early pregnancy and postdelivery estimates differed from each other in an important way. Nevertheless, the relatively small number of cases would not have obscured large changes in the point estimates of the odds ratios or a tendency for postdelivery estimates of the odds ratios to be generally higher (or lower) than their early pregnancy counterparts.

The nature of the problems of the case infants, with only seven (8.2 per cent) having severe malformations, might also have played a role in our failure to demonstrate biased reporting. If only severe malformations are associated with the phenomenon, our observations cannot provide convincing evidence about report bias. On the other hand, if maternal concern about the infant motivates (supposed) biased reporting, no matter the specific problems, then these results do make an important contribution to knowledge about biased reporting: although there was a "dose-response" increase in concern from the control through the neither-case-nor-control to the case groups, there were no corresponding differences between the postdelivery and early pregnancy reports of women in the three groups. Nor was there evidence that concern was associated with either addition or deletion when the effects of (potential) determinants of these changes were studied.

If biased reporting occurs only in response to a very strong stimulus—such as the birth of an infant with a severe malformation—then an impossibly large number of early pregnancy questionnaire respondents would be needed to demonstrate it using the design of the present study. (For an exposure with a postdelivery questionnaire prevalence in the controls of 0.05, 443 cases would be required to demonstrate biased reporting sufficient to double an early pregnancy odds ratio estimate of 1.0 (with three controls/case). In the present study, 0.54 per cent of the eligible early pregnancy questionnaire respondents gave birth to infants with severe malformations and responded to the postdelivery questionnaire. To obtain 443 cases with severe malformations would therefore have required approximately 81,000 early pregnancy questionnaire respondents.)

The lack of convincing evidence of report bias in this study does not prove that the bias does not, or cannot, exist. Theoreti-

cally, it still might be possible to demonstrate report bias if there were either many more cases similar to those in the present study or a smaller, but still fairly large, case group including only infants with severe malformations. Yet, if the bias is so small that huge samples are necessary to demonstrate it, then it is unlikely to interfere with the validity of case-control studies of the sizes usually undertaken. It would not be worth doing a larger version of the present study simply to permit changes of the magnitudes observed here to become statistically significant. Alternatively, if the nature of our cases was a limiting factor, it would be of considerable interest to study reporting by women who had all had an infant with a severe malformation. Such a study would be prohibitively expensive unless designed differently than ours.

Case-control studies of pregnancy outcome are likely to remain important tools for identifying potential hazards to reproduction. In the absence of any strong evidence of report bias, we believe that this approach is appropriate and that normal infants can properly be used as controls. However, our observation of a decreased prevalence of exposures reported by *all* women on the postdelivery questionnaire should be taken into consideration in planning these investigations. Although underreporting in the present methodological study enhanced the opportunity of demonstrating report bias, it would not be an advantage in an etiologic study where complete ascertainment of exposure is desirable to reduce symmetrical underreporting which can bias the measure of association towards the null. Furthermore, as Mitchell et al. (20) point out, the more complete the ascertainment of exposures, the less the possibility that study results can be affected by biased reporting. Thus, in case-control studies of pregnancy outcome, data collection methods must be chosen that will permit complete and accurate ascertainment of exposure.

References

1. Cannell CF, Kahn RL. Interviewing. In: Lindzey G, Aronson E, eds. The handbook of social psychology. Vol. 2. Research methods. 2nd ed. Reading, MA: Addison-Wesley, 1968:526–95.
2. Sudman S, Bradburn NM. Response effects in surveys. Chicago, IL: Adine, 1974.
3. Cannell CF, Marquis KH, Laurent A. A summary of studies of interviewing methodology. Rockville, MD: US National Center for Health Statistics, 1977. (Vital and health statistics. Series 2, Data evaluation and methods research, no. 69) (DHEW publication no. (HRA) 77–1343).
4. Baddeley A. The limitations of human memory: implications for the design of retrospective surveys. In: Moss L, Goldstein H, eds. The recall method in social surveys. London: University of London Institute of Education, 1979:13–27.
5. Bradburn NM, Sudman S. Improving interview method and questionnaire design. San Francisco: Jossey-Bass, 1981.
6. Schuman H, Scott J. Problems in the use of survey questions to measure public opinion. Science 1987;236:957–9.
7. Cole P. The evolving case-control study. J Chronic Dis 1979;32:15–27.
8. Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. Am J Med 1979;66:556–64.
9. Jick H. Recall "error" in interview studies of past drug use. (Letter.) Am J Public Health 1982; 72:405.
10. Sun M. Federal VDT study finally wins approval. Science 1986;232:1594–5.
11. Ibrahim MA, Spitzer WO. The case control study: the problem and the prospect. J Chronic Dis 1979;32:139–44.
12. Janerich DT, Polednak AP. Epidemiology of birth defects. Epidemiologic Rev 1983;5:16–37.
13. Bracken MB. Methodologic issues in the epidemiologic investigation of drug-induced congenital malformations. In: Bracken MB, ed. Perinatal epidemiology. New York: Oxford University Press, 1984:436–40, 434–5.
14. Raphael K. Recall bias: a proposal for assessment and control. Int J Epidemiol 1987;16:167–70.
15. Klemetti A, Saxén L. Prospective versus retrospective approach in the search for environmental causes of malformations. Am J Public Health 1967;57:2071–5.
16. Mantel N. Synthetic retrospective studies and related topics. Biometrics 1973;29:479–86.
17. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley and Sons, 1981.
18. Schlesselman JJ. Case-control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
20. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. Am J Epidemiol 1986;123:670–6.