THE B FILES

Case studies of bias in real life epidemiologic studies

Bias File 6. Double whammy: recall and selection bias in case-control studies of congenital malformations

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Department of Epidemiology, Biostatistics & Occupational Health

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca & jay.kaufman@mcgill.ca

Bias File 6. Double whammy: recall and selection bias in case-control studies of congenital malformations

The story

Case-control studies often rely on recall of past exposures by case and control subjects. While both cases and controls can misclassify their exposures due to poor recall (bad memory for past exposures or events), it is possible that cases and controls might have *differential* recall. That is, cases may recall differently than controls, because cases are aware that they have the disease, and controls are aware that they do not. Poor recall by both cases and controls can result in non-differential misclassification bias. On the other hand, if recall is differential (unequal) among cases and controls, then differential misclassification bias may result. Ernst Wynder, a famous epidemiologist, called this "rumination bias." Others call this "reporting bias." As Rothman and colleagues (2008) point out , "recall bias is a possibility in any case-control study that relies on subject memory, because the cases and controls are by definition people who differ with respect to their disease experience at the time of their recall, and this difference may affect recall and reporting." Coughlin's review article on recall bias in epidemiological studies is a nice overview on this topic (Coughlin SS, 1990).

Almost every textbook in epidemiology cites the classic example of case-control studies of congenital malformations. It is often stated that parents of babies with birth defects are motivated to find a cause for their child's problem, and therefore would likely to reflect back on past exposures, and more likely to report exposure relative to parents of normal live births. This differential reporting could lead to an overestimation of the odds ratio. On the other hand, parents of control children might actually underreport exposures, relative to parents of case children (presumably because they have no motivation to "ruminate"). This again could overestimate the odds ratio. So, how does one avoid this vexing problem of recall bias in studies of congenital malformations?

The controversy

Recall bias has been a topic of heated debate in the field of reproductive and perinatal epidemiology. The debate focuses on what the ideal control group should be in case-control studies of malformations. Should the control group be parents of normal children ("normal controls"), or should the control group include only parents of children with a defect other than that under study ("malformed controls")?

As is often the case, there are proponents of both approaches (Swan et al. 1992; Hook EB. 1993; Hook EB, 2000). In the past, some researchers advocated the routine use of malformed controls by suggesting that the use of normal controls will overestimate the effect (because of recall bias). The rationale for using malformed controls was to balance out the issue of selective recall by parents of malformed children. Because both case and control children will have some birth defect, it was felt that the issue of unequal or differential recall is addressed to some extent. Also, cases with other birth defects are more easily obtained than population-based controls.

Other experts argued that it was better to include normal controls because this enables direct comparison of the histories of infants affected by a selected birth defect with those without any apparent pathology. They also argued that although the use of malformed controls might appear to address the recall bias problem, two wrongs don't make a right. If cases report with bias, then finding controls who also report with bias does not necessarily fix the original bias. Also, the strategy of using malformed controls introduces a brand new problem of selection bias. Since the controls have malformations, and if the malformations in the control group were positively associated with the study exposure, then this introduces selection bias that can underestimate the odds ratio. In other words, if the study exposure was associated with the birth defects in the control group, then the exposure odds in the control group would be spuriously higher than the source population. This, in turn, would bias the odds ratio towards null, because both cases and controls may end up with fairly similar exposure histories. This is a consequence of a "teratogen nonspecificity bias". This problem would be particularly important for exposures (teratogens) that can cause a wide variety of malformations (e.g. radiation).

In addition to this selection bias introduced by using malformed controls, there is a more subtle problem. If both cases and controls have malformations, what can the odds ratio from such a case-control study tell us about the causal role of an exposure? Some argue that a case-control study that uses malformed controls cannot really tell us if the exposure is causally associated with the birth defect under study. Instead, it tells us if the exposure is more likely to be associated with a specific type of birth defect rather than other defects. In other words, the odds ratio gives us a measure of *specificity* between the exposure and a particular birth defect, but cannot tell us the causal effect of that particular exposure.

So, a solution, proposed by some researchers, is to use both types of controls. For example, Hook suggested "as the use of normal controls biases the estimate if anything high, and use of malformed controls biases the estimate if anything low, *the optimal strategy would appear to use both types of controls*... One could safely infer that the true estimate of relative risk is at least somewhere between the two, and then with more refined analysis attempt to narrow the estimate of effect." Swan and colleagues concluded that "case-control design is sensitive to both differential reporting and selection bias, and the choice of study design involves balancing these two sources of bias."

The apparent solution of using two control groups has been tried out and some studies have shown similar odds ratio estimates with healthy and malformed controls (Werler et al, 1999). While it may be reassuring if both control groups produced similar results, it is unclear as to how to proceed when the results are dissimilar with the two controls groups.

The lesson

While recall bias is a legitimate concern in case-control studies of congenital malformations, as time has gone by, there has been little evidence of widespread recall bias in case-control studies of birth defects. Investigators have learned to reduce recall bias by standardized interviews where the main exposure is one of many questions. Furthermore, there is the potential to introduce a selection bias, as noted above. Also, there is a realization that shared exposures may underlie several (apparently-disparate) types of defects. For example, Werler et al. found that multivitamins during pregnancy are associated with decreased risk of many kinds of birth defects (Werler et al 1999). To the extent these associations are real, they will be muted for any given birth defect when other birth defects are included as controls.

So, with regards to the issue of ideal control group, there is no easy solution to this problem. As Hook [2000] points out, the effects of recall bias and selection bias may go in opposite directions. Recall bias usually tends to overestimate the effect (away from the null), while selection bias will usually bias the effect towards the null. In any given study, one strategy could be better or worse than the other, as a function of the particular parameters involved. Therefore, researchers may have to try to estimate the error magnitudes and directions on a case-by-case basis, and apply sensitivity analyses based on these parameters. In general, experts in the field now agree that given the general advantages of population-based controls, and the potential problems introduced by using malformed controls, most birth defects researchers today prefer to recruit normal controls.

Sources and suggested readings*

1. Rothman KJ, Greenland S, Lash T. Modern epidemiology. 3rd Edition, Lippincott Williams & Wilkins, 2008.
2. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990;43:87-91.
3. Swan SH et al. Reporting and selection bias in case-control studies of congenital malformations. Epidemiology 1992;3:356063.
4. Hook EB. Normal or affected controls in case-control studies of congenital malformations and other birth defects: reporting bias issues. Epidemiology. 1993 Mar;4(2):182-4.
5. Hook EB. What kind of controls to use in case control studies of malformed infants: recall bias versus "teratogen nonspecificity" bias. Teratology. 2000 May;61(5):325-6.
6. Werler M et al. Multivitamin supplementation and risk of birth defects. Am J Epidemiol 1999;150:675-82.

*From this readings list, the most relevant papers are enclosed.

Acknowledgement

We are grateful to Professor Allen Wilcox (National Institute of Environmental Health Sciences (NIEHS)), and Professor Martha Werler (Boston University) for their helpful input and contribution to this case study.

Reporting and Selection Bias in Case-Control Studies of Congenital Malformations

Shanna H. Swan,[1] *Gary M. Shaw,*[2] *and Jane Schulman*[2]

Retrospective studies of congenital malformations frequently rely on exposures reported by study subjects. Differential error in exposure reporting by cases and controls, which has alternatively been referred to as "recall bias" and "reporting bias," may result in a biased effect measure. Some authors have attempted to avoid reporting bias by comparing exposures between two malformed groups, rather than between cases and nonmalformed controls. This approach, however, may introduce its own bias, which we call selection bias. Both reporting bias and selection bias are shown to be algebraically equivalent to bias arising from exposure misclassification. The magnitudes of these biases are compared for a range of plausible parametric values. The case-control design is sensitive to both differential reporting and selection bias, and the choice of study design involves balancing these two sources of bias. (Epidemiology 1992;3:356-363)

Keywords: reporting bias, selection bias, misclassification, case-control studies, congenital anomalies.

Retrospective studies of congenital malformations frequently rely on mothers' reporting of exposures. It is possible, however, that differences between case and control mothers in their concern about a prenatal exposure could affect the accuracy of exposure reporting. Bias resulting from differential error in exposure reporting by cases and controls has been alternatively referred to as "reporting bias" and "recall bias" and will be referred to here as reporting bias.

It is often hypothesized that case parents, motivated by the need to find a cause for their child's birth defect, would more likely report exposure relative to parents of normal live births. This differential overreporting would lead to an overestimate of the relative risk. A similar bias would arise if control parents "underreported" exposures relative to case parents. Alternatively, reporting bias resulting from the failure of a case mother to report an exposure has been cited as an explanation for a negative association.[1]

Reporting bias has been a subject of concern to reproductive epidemiologists for at least 30 years. In 1960, MacMahon *et al*[2] stated: ". . .if a mother's memory of events that occurred during her pregnancy or during her child's early life is stimulated by the fact that her child now has a serious disease, it will be extremely difficult to find suitable comparison individuals in whom a similar degree of stimulation could be expected. Certainly, healthy individuals will not be suitable." This belief was supported by Leck,[3] who suggested that the principal disadvantage of a retrospective study of congenital malformations was that self-reported information on prenatal exposures may be biased because a control mother may forget a particular exposure, whereas a case mother may "embroider" events that she may "blame for her misfortune."

Most studies in which this question has been examined have found little evidence for the existence of reporting bias.[4-9] Stott[10] concluded, however, that reporting bias was demonstrated when mothers of children with any malformation (not just Down syndrome, which was being studied) reported a higher frequency of stressors during pregnancy than did mothers of children without malformations. A recent study by Werler *et al*[11] also suggests reporting bias, resulting in overestimates of effect for certain exposures, but the conclusions of this study have been questioned.[12,13] On the other hand, Ahlborg[14] found that women with adverse pregnancy outcomes reported occupational exposures more accurately than controls, with controls more likely to overreport exposures, resulting in a bias toward the null.

Assessing reporting bias is difficult owing to the absence of a "gold standard" against which to compare self-reported, retrospectively collected information. Previous studies either have compared maternal responses to data available from medical or occupational records[5-7,11] or have compared responses from inter-

From the [1]Epidemiological Studies Section, California Department of Health Services, Berkeley CA, and [2]March of Dimes Birth Defects Foundation, California Birth Defects Monitoring Program, and California Department of Health Services, Emeryville, CA.

Address reprint requests to: Shanna H. Swan, Reproductive Epidemiology Program, California Department of Health Services, 2151 Berkeley Way, Annex 11, Berkeley CA 94704.

views conducted early in pregnancy with those obtained after delivery.[4,5,8,9] The first design assumes that records are more accurate than maternal recall. This assumption, which is likely to be justified only for selected medical conditions, has been questioned.[15] Moreover, abstraction of data from medical records introduces an additional source of measurement error.[16] Medical records, although incomplete, are unbiased, however, when exposure data are recorded independent of outcome. Comparing prospectively and retrospectively collected information assesses reporting consistency, but not reporting accuracy, unless the mother initially reports exposure without error.

Despite limited evidence for the existence of reporting bias, it is frequently cited as a justification for the use of malformed controls in case-control studies of congenital malformations.[11,17-20] Indeed, reporting bias could be reduced by using children with malformations as controls. This design, however, would introduce bias toward the null if the malformations included among controls were associated with the study exposure. Epidemiologic principles dictate that any disease included in the control group should be unrelated to the exposure under study.[21] This goal can be elusive, however, as Pearce and Checkoway[22] point out: "If one chooses as controls persons with a particular disease, there is almost always the chance that the 'control' disease is, in fact, related either positively or negatively to exposure in the study base, despite a lack of prior evidence for an association." Since etiologic agents are likely to increase the risk in multiple categories of malformations, depending upon the nature and timing of the exposure,[23] the possibility of an association between the study exposure and one or more control malformations is difficult to rule out.

If available, knowledge of the relevant exposure window for control malformations, together with accurate timing of the study exposure, might allow for the inclusion of control malformations unrelated to the study exposure. For example, a drug administered to combat symptoms of pregnancy (such as nausea) is unlikely to have been ingested before week 6 of pregnancy, when symptoms are first observed. In a study of such a drug, chromosomal defects—which would have occurred at (or before) conception—might be an appropriate control group for a structural malformation. In this case, the chromosomal damage would almost certainly have occurred before the exposure under study. Such clear disjunction of the exposure period and the formative period for control defects is, however, quite unusual.

Some authors who recommend the use of malformed controls recognize the potential for selection bias inherent in this design but suggest that this bias occurs only when all malformations are affected by the exposure.[24] We demonstrate that an association between the exposure under study and only a fraction of the controls can lead to appreciable bias.

We use the term "pseudo-controls" to denote subjects who are used as controls but whose malformations are associated with the study exposure. Epidemiologists are faced with a design decision: the potential for reporting bias that might result from the use of controls with normal outcomes must be balanced against the bias that may result from the inclusion of pseudo-controls in the comparison group. To assist in making this decision, we compare the magnitude of these biases for a range of plausible parametric values. We first calculate the bias in the odds ratio as a function of the proportion of exposed cases (or controls) misclassified. Two recent studies of reporting bias and accuracy[25,26] provide estimates of the frequency of exposure misclassification and the degree to which it introduces bias. Selection bias, arising from the use of pseudo-controls, is then modeled and compared with that resulting from reporting bias. It is shown that selection bias arising from the inclusion of pseudo-controls is algebraically equivalent to overreporting of exposure by controls. Both reporting bias and selection bias are shown to be a simple function of the ratio of the true odds of exposure to the observed odds of exposure in cases to that in controls.

Methods and Results

Modeling Reporting Bias

Statistical discussions of misclassification usually begin with a description of "truth" and then describe scenarios in which misclassification leads to bias. We view the problem, however, from the vantage point of the epidemiologist who, upon obtaining evidence of an association, asks, "How much is this odds ratio likely to be biased as the result of exposure misclassification?"

We consider three models for exposure misclassification. We assume, for simplicity, that exposure is dichotomous ("overreporting" refers to exposure reporting when none occurred, and conversely) and the observed (biased) odds ratio is above 1.0. We observe the following:

Observed Data:

	E	$\bar{E}$	
D	a	b	$a+b$
$\bar{D}$	c	d	$c+d$

The proportion of exposed cases that must be added to (or subtracted from) the observed two-by-two table to correct exposure misclassification will be referred to as the misclassification proportion for cases and denoted by δ_1. The corresponding proportion for controls is δ_2 ($0 \leq \delta_i \leq 1$; $i = 1,2$). In each model, we derive the true value of the odds ratio (OR_T) as a function of the observed (biased) odds ratio (OR_B), the observed prevalence of exposure in the controls (p_o), and the misclassification proportions (δ_1 and δ_2). The bias in the odds ratio (B_{OR}) is defined as:

$$B_{OR} = \frac{OR_B - OR_T}{OR_T}. \qquad (1)$$

This bias is independent of study size in all models.

First, consider Model I, in which the odds ratio has been biased upward as the result of overreporting of exposure by cases; the number of cases classified as exposed is too large. Controls are assumed to have reported exposure without error ($\delta_2 = 0$).

Model I:

	E	$\bar{E}$	
D	$a - \delta_1 a$	$b + \delta_1 a$	$a + b$
$\bar{D}$	c	d	$c + d$

Under this model, Eq 1 becomes:

$$B_{OR} = \left[\frac{\delta_1}{1 - \delta_1}\right]\left[\frac{p_o(OR_B - 1) + 1}{1 - p_o}\right], \qquad (2)$$

which is close to $\delta_1/(1 - \delta_1)$ for rare exposures. Figure 1 shows B_{OR} as a function of the misclassification proportion δ_1 for exposure prevalences of 0.05 and 0.10 and $OR_B = 2.0$.

How much bias can be expected when a moderate amount of misclassification is present? With a background exposure prevalence of 0.10 (for example, the prevalence of cigarette smoking by pregnant women), a misclassification proportion $\delta_1 = 0.20$ would result in an observed $OR_B = 2.0$, when the true odds ratio (OR_T) was actually 1.54. In this case, B_{OR} is 30%. For the range of values of δ_1 examined, the bias in the odds ratio increases with OR_B and p_o, but not rapidly. For example, when OR_B is increased to 5.0 (leaving $p_o = 0.10$ and $\delta_1 = 0.20$), the bias increases only slightly, to 38%.

For Model II, it is assumed that the odds ratio has been biased upward as a result of underreporting of exposure by controls; the number of controls classified as exposed is too small.

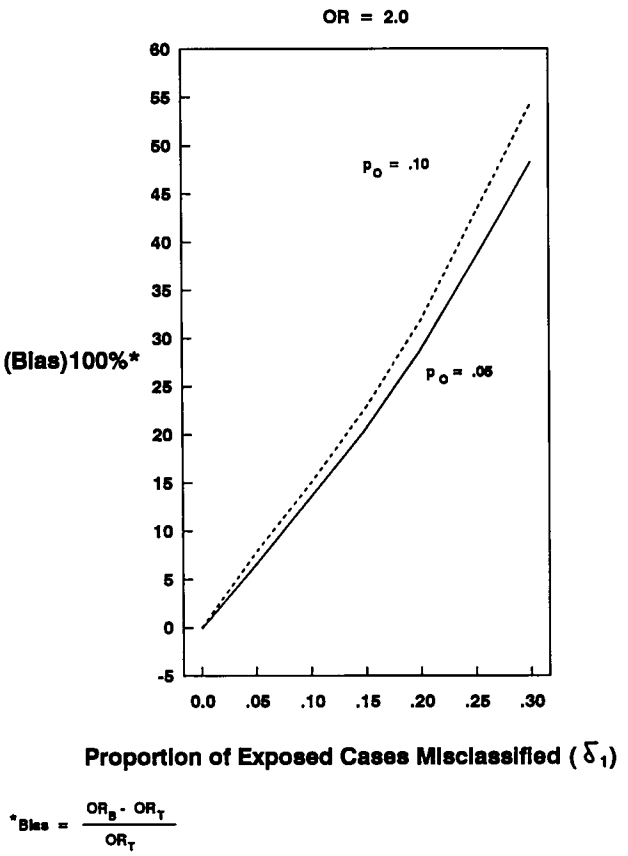

FIGURE 1. Model I: cases overreport; controls report correctly.

Model II:

	E	$\bar{E}$	
D	a	b	$a + b$
$\bar{D}$	$c + \delta_2 c$	$d - \delta_2 c$	$c + d$

In this case, the background exposure prevalence (p_o) will also be underestimated. In this model, B_{OR} takes the form:

$$B_{OR} = \frac{\delta_2}{1 - p_o(1 + \delta_2)}, \qquad (3)$$

which is independent of OR_B and is approximately equal to δ_2 for rare exposures. Figure 2 shows B_{OR} as a function of the probability of misclassification (δ_2) for two values of p_o. It can be seen that, in this model, the odds ratio is slightly less sensitive to exposure misclassification than in Model I. We compare the bias arising when $p_o = 0.10$, $OR_B = 2.0$, and $\delta_i = 0.20$ for $i = 1,2$ in Models I and II, respectively (Table 1). In this case,

the bias is 23% under Model II, compared with 30% for Model I.

In Model III, we assume underreporting of exposure by both cases and controls; the number of cases and the controls classified as exposed are both too small, and controls underreport more frequently than cases, that is, $\delta_1 < \delta_2$.

Model III:

	E	$\bar{E}$	
D	$a + \delta_1 a$	$b - \delta_1 a$	$a + b$
$\bar{D}$	$c + \delta_2 c$	$d - \delta_2 c$	$c + d$

Figure 3 shows B_{OR} as δ_1 increases from $0.25\delta_2$ to $0.75\delta_2$ for $OR_B = 2.0$. As with Model II, this model implies an error in the estimate of the background exposure prevalence (p_o). In this case:

$$B_{OR} = \frac{(1 + \delta_2)[1 - p_o(1 + OR_B\delta_1)] - (1 + \delta_1)[1 - p_o(1 + \delta_2)]}{(1 + \delta_1)[1 - p_o(1 + \delta_2)]}. \quad (4)$$

When exposure is rare, this bias is approximately $(\delta_2 - \delta_1)/(1 + \delta_1)$. In this case, B_{OR} decreases with increasing OR_B. As expected, Model III is equivalent to Model II when $\delta_1 = 0$. As the misclassification proportion for cases approaches that for controls, the bias approaches zero and may become negative. When cases and controls underreport exposure equally, classical "nondifferential" misclassification occurs, so that OR_B underestimates OR_T.[27]

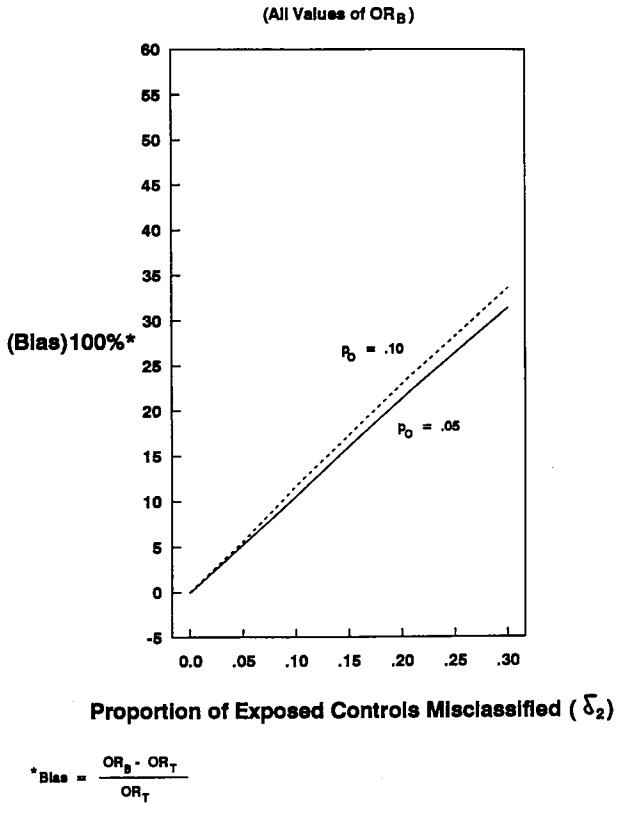

FIGURE 2. Model II: cases report correctly; controls underreport.

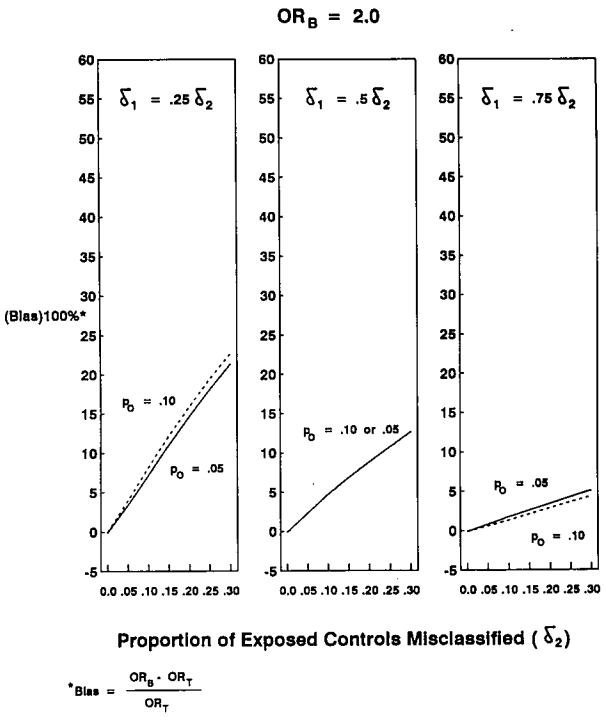

FIGURE 3. Model III: cases and controls underreport.

In Model III, both cases and controls report exposure with error. This model, which is shown below to be more likely than Model I or Model II, is the least sensitive to reporting bias. For example, when the misclassification proportion is 0.20 for controls and 0.10 for cases (still assuming $p_o = 0.10$ and $OR_B = 2.0$), the bias is only 9% (Table 1).

The amount of bias is a complex function of δ_1, δ_2, OR_B, and p_o. How large is this bias likely to be? Recent studies on reporting accuracy and consistency[25,26] provide some guidance. In the study by Drews *et al*,[25] information on 25 exposures obtained from medical records is compared with reporting of those exposures on interview by mothers of cases of sudden infant death syndrome. This study provides good estimates of reporting accuracy for exposures that are well recorded on medical records, but it is likely to overestimate misreporting for poorly recorded exposures. Fenster *et al*[26] compare responses on two interviews

TABLE 1. Comparison of True and Biased Odds Ratios as a Function of the Misclassification Proportion (Observed Prevalence of Exposure, $p_o = 0.10$)*

Model	Cases (δ_1)	Control (δ_2)	OR_B	OR_T	Bias	γ_1	γ_2	γ
I	0.20		2.0	1.54	0.30	0.77	1.00	0.77
II		0.20	2.0	1.63	0.23	1.00	1.23	0.82
III	0.05	0.20	2.0	1.73	0.16	1.06	1.23	0.87
	0.10	0.20	2.0	1.84	0.09	1.13	1.23	0.92
	0.20	0.20	2.0	2.05	−0.02	1.26	1.23	1.02
IV		0.20	2.0	2.56	−0.22	1.00	0.78	1.28

* δ_1 = proportion of exposed cases misclassified; δ_2 = proportion of exposed controls misclassified; OR_B = biased odds ratio; OR_T = true odds ratio; $\gamma_1 = (a_t/b_t)/(a_o/b_o)$; $\gamma_2 = (c_t/d_t)/(c_o/d_o)$; $\gamma = \gamma_1/\gamma_2 = OR_T/OR_B$; bias $= (1 - \gamma)/\gamma = (OR_B - OR_T)/(OR_T)$.

for nine exposures of early pregnancy. For control mothers, these interviews were prenatal and postnatal, whereas for cases of spontaneous abortion, the first interview occurred shortly after the pregnancy termination, and the second occurred several months after the birth of the control.

Estimates of misreporting from Drews *et al*[25] are consistently higher than those from Fenster *et al*,[26] as might be expected given the study designs used. In both studies, cases and controls were asked about nausea, and we compare results for this exposure. In Drews *et al*, nausea was noted on medical records for only 30% of control mothers but was reported on retrospective interview by 60% of case mothers. Reporting frequencies for nausea were also high on both first and second interview in Fenster *et al* (73% and 74%, respectively). In both studies, the estimates of reporting bias for nausea were small: 8% in Drews *et al* and 2% in Fenster *et al*.

The odds ratios based on medical record (Drews *et al*[25]) or first interview (Fenster *et al*[26]) were compared with those based on later interview. Differential exposure reporting resulted both in effect estimates biased away from the null (33% and 44% for Fenster *et al* and Drews *et al*, respectively) and in effect estimates biased toward the null (44% and 28% for Fenster *et al* and Drews *et al*, respectively). In both studies, on average, the proportion of underreporting among exposed cases and controls was greater than the proportion of overreporting. For example, in Drews *et al*, on average 36.4% of exposed cases underreported their exposures, whereas 20.4% of unexposed cases overreported exposure. Misreporting was also slightly more common in both studies among controls than cases.

The results of these studies suggest that a mixed model, with frequent reporting errors in both cases and controls, is to be expected, and the direction in which the odds ratio is altered is not predictable. These shifts in the effect measure may, therefore, reflect random variation rather than reporting bias.

MODELING SELECTION BIAS

What happens to the estimate of the odds ratio if the control group comprises a variety of malformations, some of which are related to the exposure under study? In this case, the exposure prevalence in the comparison group is overestimated owing to the inclusion of "pseudo-controls"—cases whose malformations are themselves related to the exposure under study.

To examine this bias, we again assume that a biased estimate of the odds ratio (OR_B) has been observed. Now, however, we assume that all subjects correctly report exposure. The data are again arrayed in a two-by-two table. The observed number of exposed controls (c_o) is the sum of c_t true controls and ($c_o - c_t$) pseudo-controls. Similarly, the unexposed controls include d_t true controls and ($d_o - d_t$) pseudo-controls. In this discussion, we assume that exposure is positively related to the malformations in the pseudo-controls.

Define γ as the ratio of the odds of exposure in all controls to the odds of exposure in true controls:

$$\gamma = \frac{c_o/d_o}{c_t/d_t} \qquad (5)$$

Then $OR_T = \gamma\ OR_B$, and the bias in the estimate of the odds ratio (B_{OR}) reduces to $(1 - \gamma)/\gamma$. By assumption, $\gamma \geq 1$, so that $B_{OR} \leq 0$. When no pseudo-controls are present, $\gamma = 1$, and the observed odds ratio is unbiased. As γ increases, the bias increases, in absolute value, to 1. Note that the magnitude of the bias is independent of both study size and exposure prevalence.

This scenario is algebraically equivalent to a fourth model of reporting bias, one in which cases report exposure correctly, but controls overreport exposure. Thus, bias from this source can be directly compared

with the bias occurring under the reporting bias scenarios discussed above (Table 1). For example, an estimate of $OR_B = 2.0$ would be observed when the true value was $OR_T = 2.6$ if sufficient pseudo-controls were included in the control group to produce 20% overreporting in controls.

What proportion of the control group must be "pseudo," and how strong does the association between the exposure under study and the malformation(s) among the pseudo-controls have to be to produce 20% overreporting in controls? Table 2 indicates several scenarios that would produce a bias of this magnitude. Even when only 35% of the control group consists of pseudo-controls, and the association between the malformations in the pseudo-controls and the exposure is about as strong as that between the malformations in the true controls and the exposure ($OR_T \approx OR_{pseudo}$), the odds ratio will be underestimated by 22%. As OR_T increases, a larger proportion of pseudo-controls (or a stronger association with the study exposure among pseudo-controls) will be required to achieve the same result.

These two sources of bias can be unified through γ as follows. Let a_t/b_t denote the true odds of exposure in cases, whereas a_o/b_o is the observed (biased) odds. Define γ_1 as:

$$\gamma_1 = \frac{a_t/b_t}{a_o/b_o} \tag{6}$$

and γ_2 as:

$$\gamma_2 = \frac{c_t/d_t}{c_o/d_o}, \tag{7}$$

and their ratio as γ:

$$\gamma = \frac{\gamma_1}{\gamma_2}. \tag{8}$$

TABLE 2. Alternative Control Groups Resulting in Selection Bias Equivalent to That Produced by 20% Exposure Overreporting by Controls ($\delta_2 = 0.20$)

Composition of Control Group (%)				
True Controls	Pseudo-Controls	OR_B	OR_T	OR_{pseudo}*
50	50	2.0	2.56	1.77
60	40	2.0	2.56	2.19
65	35	2.0	2.56	2.64
70	30	2.0	2.56	3.63

* OR_{pseudo} denotes the odds ratio for the association between exposure and pseudo-controls.

Simple algebra shows that, for all models, $\gamma = OR_T/OR_B$, and the bias generalizes to:

$$B_{OR} = \frac{1 - \gamma}{\gamma}. \tag{9}$$

In the particular case in which there is no misclassification among cases, $\gamma_1 = 1$, and γ reduces to:

$$\gamma = \frac{1}{\gamma_2} = \frac{c_o/d_o}{c_t/d_t}, \tag{10}$$

consistent with γ as defined above (Eq 5).

We note that the γ_i are equivalent to the inverses of the "selection odds" as defined by Kleinbaum *et al*,[28] who refer to γ_2/γ_1 as the "selection odds ratio." These quantities are illustrated in Table 1 for the specific examples discussed above.

Discussion

These results suggest that relative risk estimates from case-control studies of congenital malformations are vulnerable to both reporting bias and selection bias. The degree of robustness has been shown to vary somewhat with the model assumed. The situation described in Model I is the one most often hypothesized in the reproductive literature, although it does not appear to be the most likely. We have shown that, under this model, the bias in the odds ratio is greater than that for the other models considered. The bias increases slowly with increasing OR_B and p_o; B_{OR} is about 1.5 times the misclassification proportion among cases (δ_1) when $p_o = 0.10$ and OR_B ranges from 2 to 5. Model II is somewhat less sensitive; the bias in the odds ratio is about equal to the degree of underreporting by exposed controls (δ_2) for moderate values of the parameters, so that a 20% misclassification results in about a 20% bias in OR (still assuming that $p_o = 0.10$). In this model, the bias is independent of OR_B. When both cases and controls underreport (Model III), the bias is small and actually becomes negative as the misclassification proportion in cases approaches that for controls.

Many researchers concerned about reporting bias have attempted to guard against it by utilizing malformed controls. As we have demonstrated, the odds ratio is also affected by the bias that is likely to arise as a result of this design decision. In fact, the presence of even a small proportion of pseudo-controls, in whom the association of the exposure and disease is strong, could produce a substantial bias toward the null.

Moreover, the use of malformed controls makes interpretation of study results particularly difficult.

Although it is true that the interpretation of any study depends on the choice of reference population, the findings are dependent upon the specific malformations selected to serve as controls. Effect estimates for a disease-exposure relationship cannot be compared across studies that have used different malformation groups as controls. Bracken[24] points out that this situation is made even worse when a particular malformation is considered a case in one study and then used as a control when studying a different malformation in the same (or different) study.

Schlesselman[29] suggests that utilizing two control groups (that is, diseased and nondiseased controls) may permit both reporting bias and the effect of exposure to be assessed. It is not clear, however, how the use of two control groups allows the distinction between exposure effect and bias to be drawn. For example, consider the situation in which a study finds that the level of exposure reported by cases and diseased controls is equal (say, 20%), whereas 10% of the normal control group is exposed. Two interpretations are possible. On the one hand, there could be no association between disease and exposure, with reporting bias accounting for the elevated exposure prevalence in both cases and diseased controls. On the other hand, the odds ratio could actually be 2.0, and the exposure could be associated both with the case group and with certain malformations in the control group. Therefore, under this scenario, the use of two control groups is not likely to help in assessing reporting bias and may serve to complicate interpretation of study results.

Another approach that has been utilized to deal with potential reporting bias is to stratify the study population on the likelihood that a subject will provide biased exposure information.[30-32] Although this approach has some intuitive appeal, evidence that any population subgroup is more or less biased than any other is limited. Therefore, the stratification criteria are somewhat arbitrary. One might stratify subjects by time of interview, for example, before and after an apparently risky exposure was publicized. Alternatively, one could stratify the population on the subject's perception of the risk of exposure. For example, in a study of cardiac malformations and exposure to solvent-contaminated drinking water, women were asked how contaminated they perceived their home tapwater to be ("very," "somewhat," or "not" contaminated). Shaw *et al*[30] stratified on this perception and found that odds ratios were higher among mothers who believed that the water in their community was very contaminated. This could be taken as some evidence of reporting bias. Further studies, however, will be needed to discriminate concern due to the exposure from concern that leads to overreporting.

We have attempted to clarify some issues that must be considered when choosing a control group for case-control studies of congenital malformations. It would not be difficult to spuriously produce a small positive association as the result of reporting bias. Exposure misclassification, however, appears unlikely to produce sufficient bias to explain a strong association unless the exposure was very prevalent. It would also be easy to underestimate a positive association owing to selection bias arising from the use of malformed controls. Thus, the choice of study design involves a balancing of these two sources of bias.

Acknowledgments

The authors thank Irva Hertz-Picciotto, Kirsten Waller, and the anonymous reviewers for their helpful suggestions.

References

1. Mills JL, Rhoads GG, Simpson JL, Cunningham GC, Conley MR, Lassman MR, Walden ME, Depp OR, Hoffman HJ, and the NICHD Neural Tube Defects Study Group. The absence of a relation between the periconceptional use of vitamins and neural-tube defects. N Engl J Med 1989;321:430-435.
2. MacMahon B, Pugh TF, Ipsen J. Epidemiological Methods. Boston: Little, Brown, 1960;249-250.
3. Leck I. Backwards and forwards in search of teratogens. Early Hum Dev 1978;2/3:203-205.
4. Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. Am J Public Health 1967;57:2071-2075.
5. Zierler S, Rothman KJ. Congenital heart disease in relation to maternal use of Bendectin and other drugs in early pregnancy. N Engl J Med 1985;313:347-352.
6. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health related information: implications for retrospective studies. Am J Epidemiol 1982;116:114-122.
7. Tilley BC, Barnes AB, Bergstrath E, Labarthe D, Noller KL, Colton T, Adam E. A comparison of pregnancy history recall and medical records. Am J Epidemiol 1985;121:269-281.
8. MacKenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. Am J Epidemiol 1989;129:65-75.
9. Feldman Y, Koren G, Mattice D, Shear H, Pellegrini E, MacLeod SM. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. Teratology 1989;40:37-45.
10. Stott DH. Some psychosomatic aspects of causality in reproduction. J Psychosomatic Res 1958;3:42-55.
11. Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and non-malformed infants. Am J Epidemiol 1989;129:415-421.
12. Swan SH, Shaw GM. Re: reporting accuracy among mothers of malformed and non-malformed infants. Am J Epidemiol 1990;131:935-936.
13. Berg AT. Re: reporting accuracy among mothers of malformed and non-malformed infants. Am J Epidemiol 1990;131:936-947.
14. Ahlborg GA Jr. Validity of exposure data obtained by questionnaire. Scand J Work Environ Health 1990;16:284-248.
15. Hewson D, Bennett A. Childbirth research data: medical records

or women's reports. Am J Epidemiol 1987;125:484–491.
16. Horowitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. J Chron Dis 1984;37:825–831.
17. Mitchell AA, Rosenberg L, Shapiro S, Slone D. Birth defects related to Bendectin use in pregnancy. I. Oral clefts and cardiac defects. JAMA 1981;245:2311–2314.
18. Cordero JF, Oakley GP, Greenberg F, James LM. Is Bendectin a teratogen? JAMA 1981;245:2307–2310.
19. Rosenberg L, Mitchell AA, Shapiro S, Slone D. Selected birth defects in relation to caffeine-containing beverages. JAMA 1982;247:1429–1432.
20. Louik C, Mitchell AA, Werler MM, Hanson JW, Shapiro S. Maternal exposure to spermicides in relation to certain birth defects. N Engl J Med 1987;317:474–478.
21. Miettinen OS. Theoretical Epidemiology: Principles of Occurrence Research in Medicine. New York: John Wiley & Sons, 1985.
22. Pearce N, Checkoway H. Case-control studies using other diseases as controls: problems of excluding exposure-related diseases. Am J Epidemiol 1988;127:851–856.
23. Janerich DT, Polednak AP. Epidemiology of birth defects. Epidemiol Rev 1983;5:16–37.
24. 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;436–440.
25. Drews CD, Kraus JF, Greenland S. Recall bias in a case-control study of sudden infant death syndrome. Int J Epidemiol 1990;19:405–411.
26. Fenster L, Swan SH, Windham GC, Neutra RR. Assessment of reporting consistency in a case-control study of spontaneous abortions. Am J Epidemiol 1991;133:477–488.
27. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 1977;105:488–495.
28. Kleinbaum DG, Morgenstern H, Kupper L. Selection bias in epidemiologic studies. Am J Epidemiol 1981;113:452–463.
29. Schlesselman JJ. Case-Control Studies. Oxford: Oxford University Press, 1982.
30. Shaw GM, Swan SH, Harris JA, Malcoe LH. Maternal water consumption during pregnancy and congenital cardiac anomalies. Epidemiology 1990;1:206–211.
31. Ozonoff D, Colten ME, Cupples A, Heeren T, Schatzkin A, Mangione T, Dresner M, Colton T. Health problems reported by residents of a neighborhood contaminated by a hazardous waste facility. Am J Ind Med 1987;11:581–597.
32. Roht LH, Vernon SW, Weir FW, Pier SM, Sullivan P, Reed LJ. Community exposure to hazardous waste disposal sites: assessing reporting bias. Am J Epidemiol 1985;122:418–433.

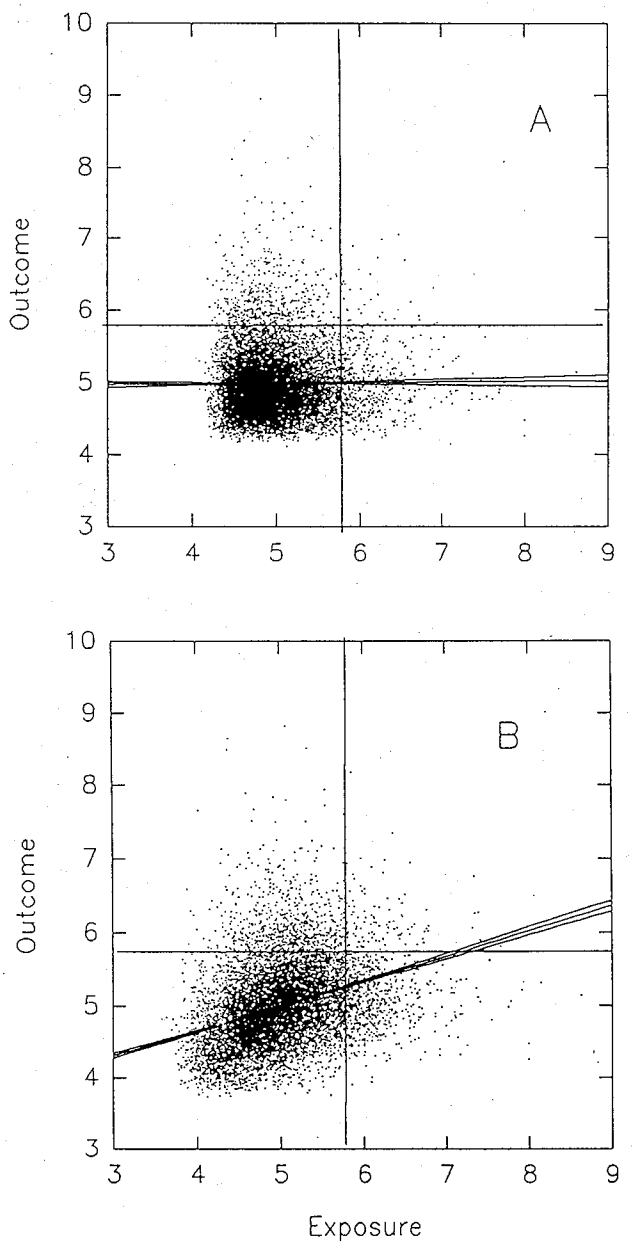

FIGURE 1. A, a random sample of 10,000 data points for exposure and outcome, both log normally distributed (mean = 5.0, standard deviation = 0.45). Regression coefficient (β = 0.05) with 95% confidence limits is indicated, and the 95th percentiles for both variables are indicated by the *horizontal* and *vertical lines*.
B, data from Figure 1A subject to error. 2,500 random data points have false low values, and 2,500 data points have false high values, for both exposure and outcome (error for both variables = 0.45). The regression coefficient = 0.34

underreporting will prove more important when the variables are continuous or polytomous. Errors that are absolute values independent of the magnitude of the true value will also influence the sensitivities more than the specificities when categorized, which will tend to increase the relative influence of underreporting. In the example with underlying continuous variables (Figure 1*B*), the influence from underreporting is substantial and of the same magnitude as the influence from overreporting concerning the regression coefficient. The 2,500 observation units with false low values alone would give a regression coefficient of 0.16, and the odds ratio for the dichotomized variables would be 1.29. The effect from the 2,500 false high values alone would yield a regression coefficient of 0.16 and an odds ratio of 1.76.

Petter Kristensen

National Institute of Occupational Health,
Gydas vei 8, Pb. 8149, Dep. N-0033,
Oslo 1, Norway (address for correspondence)

References

1. Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. Epidemiology 1992;3:210–215.
2. Chavance M, Dellatolas G, Lellouch J. Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relation between handedness and immune disorders. Int J Epidemiol 1992;21:537–546.

Normal or Affected Controls in Case-Control Studies of Congenital Malformations and Other Birth Defects: Reporting Bias Issues

To the Editor:

There has been extensive ongoing discussion of the question of whether to use affected *or* unaffected controls in case-control studies of congenital malformations and other birth defects. The paper by Swan *et al*[1] cites numerous studies on both sides of the issue but concludes lamely that there's no best answer and that the choice of study design involves balancing two sources of bias: reporting bias and what these authors call "selection bias." Reporting bias results from presumed underreports of exposure by normal controls and is a bias to falsely high relative risk. Selection bias results from presumed effects of exposure upon some of the outcomes among the malformed or otherwise defective controls and is a bias to falsely low relative risk.

Why not use *both* sets of controls?[2] Swan *et al* in fact note that Schlesselman[3] in essence suggested this some years ago. Swan *et al* object to this procedure, however, on what I believe to be the spurious grounds that "it is not clear how the use of the two control groups allows the distinction between exposure effect and [reporting] bias to be drawn." They offer an implausible scenario in which the exposure proportion is 20% both in cases and in malformed controls and only 10% in the normal controls. They note correctly that this result could be due just to underreporting among controls. But they suggest that this difference could also result just as well if exposure was associated with both the case malformation and *certain* malformations in the control group. I disagree. If the exposure is associated with only *certain* of the malformed controls, then while one may still expect an excess exposure level among the malformed controls, an effect *equal* in controls to that in the cases is not plausible. I know of no teratogen that causes equal rates of all malformations. By using controls with a *wide spectrum* of different malformations, one avoids these problems. Of course, Swan *et al* might reply that there may be a certain few specific malformations in the controls that have an even greater association with exposure than does the malformation under study. But, of course, one must assume that the investigator uses some common sense and does not, for example, include any cases with limb reduction defects as malformed controls in a case-control study of thalidomide and ear defects! The investigator must also examine the observed *pattern* of association within the malformed control group, for he or she might even discover a new association thereby.

Even these considerations still do not address a main goal of case-control studies, which is to define an effect estimate. If use of one estimator *may* produce an overestimate of effect and another estimator *may* produce an underestimate of effect, it appears only common sense to use both and see what the results are. They may turn out to be only trivially different,[2] in which case the investigator can go on to more productive issues. If they are substantially different, then the investigator can go on to explore the reasons for the difference, for example by determining whether some specific malformation is increased among the malformed controls. Indeed, if the lower limit of the range indicates a positive association, that may be sufficient for initial investigation. Thus, the use of both sets of controls may often be sufficient, practically speaking, to avoid the theoretical objections raised by Swan *et al*.

Ernest B. Hook

School of Public Health, University of California
Berkeley, CA 94720 (address for correspondence)

References

1. Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case-control studies of congenital malformations. Epidemiology 1992;3:356-363.
2. Hook EB, Cross PK. Cigarette smoking and Down syndrome. Am J Hum Genet 1985;37:1216-1224.
3. Schlesselman JJ. Case-Control Studies. Oxford: Oxford University Press, 1982.

The Authors Reply:

We welcome the comments of Dr. Hook concerning our recent publication on reporting and selection bias in studies of congenital malformations.[1] Hook's critique relies heavily on a study that utilized two groups of controls, one comprised of "normal" offspring and one the "defect" control group.[2] This study, by Hook and Cross, examining the association between maternal smoking and Down syndrome, nicely illustrates the bias inherent in using a control group including malformed infants. The odds ratios for maternal smoking around the time of conception were 0.56 and 0.58 when using normal and defect controls, respectively. It is because of the similarity of these estimates that Hook used this example in his letter. With respect to any history of maternal smoking, the situation is different, however, and this difference is not mentioned by Hook in the article, or the letter. When using the normal control group (adjusting for maternal age), maternal smoking was associated with a decreased risk (Mantel-Haenszel, odds ratio = 0.72). If the comparison is made to the defect controls, however, the corresponding estimate is 1.16. Thus, using defect controls results in a bias in the estimate of 61%. This bias is larger than any suggested by us in our paper, and it counters Hook's criticism that the scenarios proposed by us were "implausible."

In fact, because the etiology of most malformations is unknown, one has little evidence to support the contention that the scenarios we propose are implausible. Etiologic agents are likely to increase risk in multiple categories of malformations, depending on the nature and timing of exposure. Therefore, we argued that a control group including malformations has a likelihood of including defects associated with

the exposure under study. Contrary to Hook's comment, we made no assumption that the exposure under study increased the risk *equally* in all malformation categories, and we agree that such a scenario would, indeed, be implausible.

We agree with Hook that the investigator should use "common sense" and consult the literature. But even the most diligent investigator will find it difficult to select, *a priori*, a sizable group of malformations for which the possibility of an association with the exposure under study can be ruled out.

Shanna H. Swan

Reproductive Epidemiology Section,
California Department of Health Service,
2151 Berkeley Way, Annex 11, Berkeley, CA 94704
(address for correspondence)

Gary M. Shaw

March of Dimes Birth Defects Program,
California Birth Defects Monitoring Program, Emeryville, CA

Jane Schulman

Battelle Memorial Institute, Arlington, VA

References

1. Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case-control studies of congenital malformations. Epidemiology 1992;3:356–363.
2. Hook EB, Cross PK. Cigarette smoking and Down syndrome. Am J Hum Genet 1985;37:1216–1224.

Exact Stratification of Person-Years

To the Editor:

Macaluso[1] describes a computer program for the exact stratification of person-years. His article gives the misleading impression that "exact" (that is, correct) allocation of person-years to the appropriate categories (strata) of time-related variables is not commonly done by existing programs. The article also states that current programs either use approximate methods to calculate person-years or methods that have not been published, and it further suggests that existing programs that do allow for exact allocation of person-years are expensive on mainframes and difficult to implement on personal computers.

Certainly, we agree that person-time must be allocated "exactly," or else misclassification of person-time by category of time-related variables (such as age or duration of exposure) could occur. To do this, any computer program must calculate specific cutpoints in time for time-related variables, for each subject as he or she is followed over time. All eligible person-time must then be allocated to the appropriate categories of the time-related variables in relation to these cutpoints. Specific algorithms to do this may vary and may be more or less efficient. The National Institute for Occupational Safety and Health (NIOSH) Life Table Analysis System (LTAS)[2] allocates each person-day to the time-related variables age, calendar time, duration of exposure, and time-since-first-exposure. This allocation is done using an algorithm similar to the method D described by Macaluso.[1] This allocation is certainly exact. It can be expensive on mainframe computers for large cohorts. It is quite feasible, however, on personal computers. We are in the process of converting the NIOSH LTAS for use on a personal computer. Initial test runs have shown that a file of 2,500 workers, with 14,000 job histories, 500 deaths, and 67,000 person-years, can be processed in 1 minute.

Although we are ignorant of the internal details of other programs to generate person-time, such as OCMAP,[3] PERSON-YEARS,[4] and the Monson[5] program (all of which run on personal computers), we assume that they work on much the same principles as our own. Although Macaluso is to be commended for writing another computer program to calculate person-time, we are inclined to believe that his program is not unique and that other existing programs are also "exact."

Kyle Steenland
Steve Spaeth
Alberto Salván

National Institute for Occupational Safety and Health,
4676 Columbia Parkway, R-15, Cincinnati, OH 45226-1998
(address for correspondence)

References

1. Macaluso M. Exact stratification of person-years. Epidemiology 1992;3:441–448.
2. Steenland K, Beaumont J, Spaeth S, Brown D, Okun A, Jurcenko L, Ryan B, Phillips S, Roscoe R, Stayner L, Morris J. New developments in the NIOSH life table analysis system. J Occup Med 1990;32:1091–1098.
3. Marsh GM, Preininger M. OCMAP: a user-oriented occupational cohort mortality analysis program. Am Stat 1980;34:245–246.
4. Coleman M, Douglas A, Hermon C, Peto J. Cohort study analysis with a FORTRAN computer program. Int J Epidemiol 1986;15:134–137.
5. Monson RR. Analysis of relative survival and proportional mortality. Comput Biomed Res 1974;7:325–332.

TERATOLOGY 61:325–326 (2000)

Letters to the Editor

What Kind of Controls to Use in Case Control Studies of Malformed Infants: Recall Bias Versus "Teratogen Nonspecificity" Bias

To the Editor:

Many investigations of the causes of a specific birth defect use retrospective case control studies of affected livebirths. Such studies maximize efficiency and minimize cost compared with prospective studies of pregnant or pre-pregnant women who are followed to delivery or termination of pregnancy. But they entail some indirect cost. This is especially the case for a study that investigates these variables retrospectively, that is, after investigators ascertain the defect and parents are aware of it.

Investigators have debated two possible strategies to choice of controls in such retrospective studies: one uses (parents of) malformed controls, the other (parents of) normal controls. The use of normal controls enables direct comparison of the histories of infants affected by a selected defect with those of infants without any apparent pathology. But concern about selective memory of exposure or other variables by parents of affected infants, that is, "recall bias" potentially undermines interpretation of any positive association in such a study. That is, the use of normal controls may be expected a priori to lead if anything to an overestimate of an effect. Certainly, many, but by no means all investigations of such have found little evidence for such a bias (see references in Swan et al., '92). But observations from one population studied at one time may not extrapolate readily to another. One can never exclude such a possibility in a particular retrospective study using normal controls only with some ad hoc investigation.

For this reason, many epidemiological investigators of human defects use routinely only parents of malformed controls, that is, those with a defect other than that under direct investigation. One hopes that any selective recall by parents will "balance out" in comparing cases and malformed controls. If, however, a particular exposure can also cause other malformations represented in the malformed controls, then the estimate of effect will be biased toward no effect. One may call this a consequence of a "teratogen nonspecificity" bias, which leads to an underestimate of effect.

Certainly most teratogenic "exposures" are relatively specific in their effects—with the possible exception of diabetes—so that the effect of the latter bias if any may be only trivial. But Prieto and Martínez-Frías ('99) neatly demonstrate how, in the estimation of effect of maternal valproic acid exposure and spina bifida, such bias from the use of just malformed controls can mislead. Their data indicate a threefold difference in estimation of effect if normal (odds ratio about 50) or malformed controls (odds ratio about 15) were used for spina bifida cases. (The higher figure implies more than a doubling in the estimated absolute risk for a child with spina bifida after maternal exposure, from the 1–2% figure in the literature to 3–4%.).

Note that the effects of recall bias and teratogenicity non-specific bias are in opposite directions. In any particular study, one cannot predict a priori which of the two are present or are of greater magnitude. But as use of normal controls biases the estimate if anything high, and use of malformed controls biases the estimate if anything low, *the optimal strategy would appear to be to use both types of controls.* Schlesselman ('82) in essence suggested this some years ago, yet surprisingly this common sense strategy has not only been challenged (Swan et al., '92, '93) but has not been widely used (see also Hook, '92, '93 for comment). One could safely infer that the true estimate of relative risk is at least somewhere between the two, and then with more refined analysis attempt to narrow the estimate of effect.

I emphasize, however, that all such approaches suffer the defect of any "observational" study—even those

*Correspondence to: Ernest B. Hook, School of Public Health, Warren Hall, MCH-M.C. 7360, University of California, Berkeley, CA 94720-7360. E-mail: ebhook@socrates.berkeley.edu

Received 10 November 1999; Accepted 2 December 1999

that are prospective non-case control. In the absence of random or at least nondifferential assignment of exposure, the possibility of undetected confounding due to association of exposure or defect with some uncontrolled variable complicates causal interpretation of the association of a birth defect and maternal exposure or other parental variables concerning the pregnancy.

ERNEST B. HOOK*
School of Public Health
University of California, Berkeley
Berkeley, California

Department of Pediatrics
University of California, San Francisco
San Francisco, California

LITERATURE CITED

Hook EB. 1993a. Normal and/or affected controls in case-control studies of congenital malformation and other birth defects. Epidemiology 4:182–183.

Hook EB. 1993b. Estimation and the nature of controls in birth defect studies. Epidemiology 4:558–559.

Príeto L, Martínez-Frías ML. 1999 Case-control studies using only malformed infants: are we interpreting the results correctly? Teratology 60:1–2.

Schlesselman JJ. 1982. Case control studies: design, conduct, analysis. New York: Oxford University Press. p 135–136.

Swan SH, Shaw GM, Schulman J. 1992. Reporting and selection bias in case-control studies of congenital malformations. Epidemiology 3:356–363.

Swan SH, Shaw GM, Schulman J. 1993. Normal and/or affected controls in case-control studies of congenital malformation and other birth defects. Epidemiology 4:183–184.

American Journal of Epidemiology

Vol. 150, No. 7
Printed in U.S.A.

Multivitamin Supplementation and Risk of Birth Defects

Martha M. Werler,[1] Catherine Hayes,[2] Carol Louik,[1] Samuel Shapiro,[1] and Allen A. Mitchell[1]

It is widely accepted that supplementation with folic acid, a B vitamin, reduces the risk of neural tube defects (NTDs). This case-control study tested the hypothesis that multivitamins reduce risks of selected birth defects other than NTDs. Infants with and without birth defects and aborted fetuses with birth defects were ascertained in the greater metropolitan areas of Boston, Philadelphia, and Toronto during 1993–1996. Mothers were interviewed within 6 months after delivery about a variety of factors, including details on vitamin use. Eight case groups were included: cleft lip with or without cleft palate, cleft palate only, conotruncal defects, ventricular septal defects, urinary tract defects, limb reduction defects, congenital hydrocephaly, and pyloric stenosis (*n*'s ranged from 31 to 186). Controls were 521 infants without birth defects (nonmalformed controls) and 442 infants with defects other than those of the cases (malformed controls). Daily multivitamin supplementation was evaluated according to gestational timing categories, including periconceptional use (28 days before through 28 days after the last menstrual period). Odds ratios (ORs) below 1.0 were observed for all case groups except cardiac defects, regardless of control type. For periconceptional use, ORs with 95% confidence intervals that excluded 1.0 were estimated for limb reduction defects using both nonmalformed controls (OR = 0.3) and malformed controls (OR = 0.2) and for urinary tract defects using both nonmalformed controls (OR = 0.6) and malformed controls (OR = 0.5). Statistically significant ORs for use that began after the periconceptional period were observed for cleft palate only and urinary tract defects. These data support the hypothesis that periconceptional vitamin supplementation may extend benefits beyond a reduction in NTD risk. However, other than folic acid's protecting against NTDs, it is not clear what nutrient or combination of nutrients might affect risk of other specific defects. *Am J Epidemiol* 1999;150:675–82.

abnormalities; pregnancy; teratogens; vitamins

There has been long-standing interest in the relation between vitamin supplementation and the risk of birth defects. In particular, the well documented reduction in neural tube defect risk induced by folic acid has prompted widespread health advisories promoting daily supplementation among all women of childbearing age (1, 2). Recently, reports have also suggested that multivitamin supplementation before pregnancy or early in pregnancy reduces the risks of other specific congenital malformations, including defects of the lip and palate (3, 4), heart (5–8), limbs (5, 7, 9), urinary tract (5, 10), brain (5), and pylorus muscle (5). The present study tested the hypothesis that periconceptional multivitamin supplementation reduces the risks of these specific birth defects, using data collected in a large case-control study.

Received for publication May 26, 1998, and accepted for publication March 4, 1999.

Abbreviations: CI, confidence interval; CL/P, cleft lip with or without cleft palate; LM, lunar month.

[1] Slone Epidemiology Unit, School of Public Health, Boston University, Boston, MA.
[2] School of Dental Medicine, Harvard University, Boston, MA.
Reprint requests to Dr. Martha M. Werler, Slone Epidemiology Unit, 1371 Beacon Street, Brookline, MA 02446.

MATERIALS AND METHODS

The data were collected by the Boston University Slone Epidemiology Unit Birth Defects Study in the greater metropolitan areas of Boston, Massachusetts; Philadelphia, Pennsylvania; and Toronto, Ontario, Canada (11). Infants with major malformations identified by 5 months of age were ascertained in birth hospitals and in tertiary care hospitals, as were women whose pregnancies had been terminated because of a malformed fetus. Beginning in 1993, a random sample of nonmalformed infants was also ascertained from birth hospitals. Because of staffing limitations, not all ascertained subjects were approached for interview. Each month, interview subjects were selected to include: 1) those with any of approximately 10 "priority" diagnoses (a list that reflected the then-current research interests of the program); 2) an approximate 25 percent random sample of ascertained nonmalformed subjects; and 3) subjects with malformations other than the "priority" diagnoses who resided in the same general geographic area as subjects selected under points 1 and 2. Because interviews were conducted in person, most often in the subject's home, the

third criterion served to maximize interview efficiency by minimizing travel time.

The questionnaire was administered by a study nurse no more than 6 months after delivery. Questions were asked about demographic, reproductive, and medical factors; medication, alcohol, and cigarette use; and dietary intake. The product name, starting and stopping dates, and frequency of use were recorded for each vitamin supplement that was taken at any time from 2 months before the last menstrual period through the end of the pregnancy. The present analysis includes data from interviews conducted between 1993 and 1996.

For the present study, eight case groups were identified based on defects that had been reported in the literature (3–10) as possibly being inversely associated with multivitamin use (table 1). The case groups excluded subjects with any neural tube defect, known chromosomal anomaly, or Mendelian inherited disorder. Of the case defect groups, only conotruncal defects and limb reduction defects were considered Birth Defect Study "priority" diagnoses for interview selection purposes as described above. Controls were 521 infants with no major structural malformations (nonmalformed controls). We also created a secondary control group comprising the 442 subjects with major malformations, after excluding infants with neural tube defects and the eight case defects, to address the possibility of recall bias (malformed controls). The distribution of malformations in the malformed control group was as follows: genital defects, 16 percent; talipes varus or valgus defects, 14 percent; diaphragmatic hernia, 10 percent; gastroschisis, 9 percent; craniosynostosis, 7 percent; intestinal atresia, 6 percent; and various other defects, 38 percent. Gastroschisis (1993–1994) and craniosynostosis were Birth Defect Study "priority" diagnoses for interview selection purposes as described above.

TABLE 1. Case groups* in the Slone Epidemiology Unit Birth Defects Study, 1993–1996

Diagnosis	No.
Cleft lip with or without cleft palate	114
Cleft palate only	46
Conotruncal defects	157
Transposition of the great arteries	63
Tetralogy of Fallot	49
Common truncus, aorticopulmonary window, double-outlet right ventricle, pulmonary artery atresia/stenosis with ventricular septal defect, subarterial ventricular septal defect, interrupted aortic arch	45
Any ventricular septal defect	186
Limb reduction defects, including reduction of the transverse, postaxial, preaxial, and intercalary types	31
Urinary tract defects, including defects of the kidney, ureter, bladder, and urethra	184
Obstructive urinary tract defect	87
Congenital hydrocephaly	44
Pyloric stenosis	60

* Infants with known Mendelian inherited disorders, chromosomal anomalies, or neural tube defects were excluded.

Mothers who resided within our geographically defined catchment areas (approximately a 2-hour drive from either Boston, Philadelphia, or Toronto), who spoke English, and whose physicians provided consent for us to contact them were eligible for inclusion. The physicians of 7 percent of malformed cases, 3 percent of nonmalformed controls, and 8 percent of malformed controls refused participation. Because more subjects were ascertained than could be interviewed and because we set a limit that interviews had to be completed within 6 months after delivery, the mothers of 55 percent of ascertained cases, 72 percent of ascertained nonmalformed controls, and 57 percent of ascertained malformed control subjects were not asked to participate. Of those who were asked to participate, the mothers of 66 percent of cases, 65 percent of nonmalformed controls, and 66 percent of malformed controls agreed to be interviewed.

Multivitamin supplementation was defined as daily use of a supplement that contained at least two water-soluble vitamins and at least two fat-soluble vitamins. Supplementation was categorized by the beginning of first use, according to lunar month of pregnancy (28-day months beginning with the last menstrual period), as follows: the month before the last menstrual period through lunar month 1 (pre-LM1); lunar month 2 (LM2); lunar month 3 (LM3); and lunar month 4 (LM4). For each case group, odds ratios were estimated for developmentally appropriate gestational timing categories of supplementation. Specifically, conotruncal defects and ventricular septal defects develop before mid-LM2, so odds ratios were estimated for the pre-LM1 and LM2 categories, with no use during those time periods designated as the reference categories. Cleft lip with or without cleft palate (CL/P) and limb reduction defects develop by the end of LM3, so odds ratios were estimated for pre-LM1, LM2, and LM3, with no use in those time periods being defined as the reference categories. For the defects for which developmental timing is either not known (pyloric stenosis), varies across specific defects within the group (urinary tract defects), or occurs later in gestation (cleft palate only, hydrocephaly), pre-LM1, LM2, LM3, and LM4 were examined, with no

use in those time periods being defined as the reference categories. In addition, an "etiologically relevant" summary measure was estimated for each defect by combining the appropriate timing categories (e.g., for CL/P, use beginning at any time between pre-LM1 and LM3, and for cleft palate alone, use beginning at any time between pre-LM1 and LM4). Odds ratios were not estimated if there were fewer than four exposed subjects in a timing category.

Multivariate-adjusted odds ratios and 95 percent confidence intervals were estimated using unconditional logistic regression models (12). Factors found to be related to multivitamin use were included in the multivariate models: maternal age (<20, 20–24, 25–29, and ≥30 years), maternal education (<12, 12, 13–15, and ≥16 years), maternal race (White/Nonwhite), planned pregnancy (yes/no), nausea and vomiting during the first lunar month of pregnancy (yes/no), and geographic center (Boston, Philadelphia, and Toronto).

RESULTS

To assess possible demographic differences among case and control participants and nonparticipants, we examined community-level median family income. Zip code information for US mothers was linked to 1990 US Census data on median family income (13). The distributions of income categories (<$25,000, $25,000–$34,999, $35,000–$44,999, $45,000–$54,999, $55,000–$64,999, and ≥$65,000 per year) were similar (data not shown) for interviewed and noninterviewed mothers of cases, nonmalformed controls, and malformed controls, with one exception. Among interviewed mothers, the lowest category of zip code-linked income (<$25,000) was prevalent in 3 percent, 2 percent, and 4 percent of cases, nonmalformed controls, and malformed controls, respectively. The corresponding prevalences for noninterviewed mothers were 9 percent, 9 percent, and 10 percent, respectively.

The distribution of supplement use according to gestational timing is shown in table 2 for mothers of cases and controls. The prevalence of no use in the first 4 lunar months of pregnancy ranged from 11 percent to 26 percent. Across all case and control groups, the majority of women began supplementation during pre-LM1 or LM2, followed by declines in the prevalence of first use during LM3 and LM4. Women who used supplements less than daily did so in a variety of patterns across the early months of gestation, but few were less-than-daily users throughout the 5-lunar-month period under study. For example, some women began using prenatal vitamins during the first months of pregnancy, but took them infrequently because of nausea; they then became daily users later in the first trimester when the nausea had subsided. Because our analysis attempted to examine the effects of gestational timing by month of first use, it was difficult to categorize the erratic patterns of less-than-daily users. Therefore, they were excluded from risk estimation.

Odds ratio estimates (and 95 percent confidence intervals) are presented in table 3 for case groups, with nonmalformed controls used as the reference group. Risk estimates below 1.0 were observed for all case groups except cardiac defects. Statistically significant odds ratios were observed for cleft palate only and first use in LM2; for limb reduction defects and first use in

TABLE 2. Daily multivitamin supplementation according to gestational timing* among mothers of cases and controls, Slone Epidemiology Unit Birth Defects Study, 1993–1996

Subjects	Supplement use											
	None		Lunar month (LM)									
			Pre-LM1		LM2		LM3		LM4		Less than daily pre-LM4	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Cases												
Cleft lip with or without cleft palate (n = 114)	20	18	26	23	32	28	18	16	9	8	9	8
Cleft palate only (n = 46)	11	24	15	33	4	9	7	15	3	7	6	13
Conotruncal defect (n = 157)	19	12	40	26	42	27	24	15	10	6	22	14
Ventricular septal defect (n = 186)	20	11	52	28	52	28	24	13	15	8	23	12
Limb reduction defect (n = 31)	8	26	4	13	8	26	6	19	2	7	3	10
Urinary tract defect (n = 184)	37	20	43	23	36	20	32	17	12	7	24	13
Hydrocephaly (n = 44)	7	16	9	21	14	32	6	14	4	9	4	9
Pyloric stenosis (n = 60)	12	20	14	23	22	37	5	8	2	3	5	8
Controls												
Nonmalformed (n = 521)	64	12	140	27	141	27	85	16	31	6	60	12
Malformed (n = 442)	58	13	128	29	114	26	70	16	20	5	52	12

* Lunar month of pregnancy (see "Materials and Methods").

TABLE 3. Multivariate odds ratios estimated for timing of daily multivitamin supplementation and selected birth defects using nonmalformed controls, Slone Epidemiology Unit Birth Defects Study, 1993–1996

Case group	Timing of supplement use									
	Lunar month (LM)								"Etiologically relevant"*	
	Pre-LM1		LM2		LM3		LM4			
	OR†	95% CI†	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Cleft lip with or without cleft palate‡	0.7	0.4, 1.3	0.8	0.5, 1.5	0.7	0.4, 1.4		—§	0.8	0.5, 1.3
Cleft palate only¶	0.5	0.2, 1.2	0.1	0.04, 0.4	0.5	0.2, 1.4		—	0.4	0.2, 0.8
Conotruncal defect**	1.0	0.6, 1.6	1.0	0.6, 1.7				—	1.0	0.7, 1.5
Ventricular septal defect**	1.2	0.7, 1.9	1.2	0.8, 1.9				—	1.2	0.8, 1.8
Limb reduction defect‡	0.3	0.1, 0.9	0.6	0.2, 1.5	0.6	0.2, 1.8		—	0.5	0.2, 1.1
Urinary tract defect¶	0.6	0.3, 1.0	0.4	0.3, 0.8	0.7	0.4, 1.3	0.7	0.3, 1.5	0.6	0.4, 0.9
Hydrocephaly¶	0.7	0.2, 2.0	1.1	0.4, 2.9	0.8	0.2, 2.6	0.8	0.2, 3.2	0.8	0.3, 2.1
Pyloric stenosis¶	0.7	0.3, 1.6	1.0	0.4, 2.1	0.3	0.1, 1.0		—	0.7	0.3, 1.4

* Pre-LM2 for conotruncal and ventricular septal defects; pre-LM3 for cleft lip with or without cleft palate and limb reduction defects; pre-LM4 for cleft palate only, urinary tract defect, hydrocephaly, and pyloric stenosis.
† OR, odds ratio; CI, confidence interval.
‡ Reference category: no use or use that began after LM3.
§ Fewer than four exposed cases.
¶ Reference category: no use or use that began after LM4.
** Reference category: no use or use that began after LM2.

pre-LM1; and for urinary tract defects and first use in LM2. In contrast, odds ratios for both conotruncal defects and ventricular septal defects were close to 1.0. Conotruncal defects were further divided into specific defects (data not shown): Among 49 cases with tetralogy of Fallot, the odds ratio was 1.7 (95 percent confidence interval (CI): 0.8, 3.6) for pre-LM2 supplementation; among 63 cases with transposition of the great arteries, the corresponding estimate was 0.9 (95 percent CI: 0.5, 1.7). When obstructive urinary tract defects were examined (data not shown), the odds ratio was 0.5 (95 percent CI: 0.3, 1.0) for pre-LM4.

The possibility of differential maternal recall of multivitamin use between malformed and nonmalformed subjects prompted us to also estimate risks using malformed controls. Table 4 presents odds ratio estimates for each case group and the relevant supplement timing categories. Estimates were similar to those derived using the nonmalformed control group (table 3). For tetralogy of Fallot, the odds ratio was 1.4 (95 percent CI: 0.7, 2.8) for pre-LM2 use; the corresponding estimate for transposition of the great arteries was 0.9 (95 percent CI: 0.5, 1.6). The obstructive urinary tract defect risk estimate for pre-LM4 use was 0.5 (95 percent CI: 0.3, 0.9).

Twenty-one case women and eight malformed control women, but no nonmalformed control women, had undergone a pregnancy termination. Because these women were interviewed approximately 4 months

TABLE 4. Multivariate odds ratios estimated for timing of daily multivitamin supplementation and selected birth defects using malformed controls, Slone Epidemiology Unit Birth Defects Study, 1993–1996

Case group	Timing of supplement use									
	Lunar month (LM)								"Etiolotgically relevant"*	
	Pre-LM1		LM2		LM3		LM4			
	OR†	95% CI†	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Cleft lip with or without cleft palate‡	0.5	0.3, 1.0	0.8	0.4, 1.4	0.8	0.4, 1.5		—§	0.7	0.4, 1.1
Cleft palate only¶	0.4	0.2, 1.1	0.2	0.1, 0.6	0.5	0.2, 1.4		—	0.4	0.2, 0.9
Conotruncal defect**	0.8	0.5, 1.3	1.0	0.6, 1.7				—	0.9	0.6, 1.4
Ventricular septal defect**	1.0	0.6, 1.5	1.2	0.8, 1.9				—	1.1	0.7, 1.6
Limb reduction defect‡	0.2	0.1, 0.7	0.5	0.2, 1.4	0.7	0.2, 2.1		—	0.4	0.2, 1.0
Urinary tract defect¶	0.5	0.3, 0.8	0.5	0.3, 0.8	0.7	0.4, 1.2	1.0	0.4, 2.5	0.5	0.3, 0.9
Hydrocephaly¶	0.6	0.2, 1.8	1.0	0.4, 2.7	0.7	0.2, 2.3	1.5	0.4, 6.1	0.8	0.3, 2.1
Pyloric stenosis¶	0.7	0.3, 1.6	1.1	0.5, 2.4	0.4	0.1, 1.1			0.7	0.3, 1.5

* Pre-LM2 for conotruncal and ventricular septal defects; pre-LM3 for cleft lip with or without cleft palate and limb reduction defects; pre-LM4 for cleft palate only, urinary tract defect, hydrocephaly, and pyloric stenosis.
† OR, odds ratio; CI, confidence interval.
‡ Reference category: no use or use that began after LM3.
§ Fewer than four exposed cases.
¶ Reference category: no use or use that began after LM4.
** Reference category: no use or use that began after LM2.

closer to the time of conception than were women who delivered near term, and because the diagnoses in approximately 60 percent of the terminations were not confirmed by autopsy, we reestimated odds ratios after excluding these subjects. There were no appreciable changes in risk estimates (data not shown).

DISCUSSION

In this study, we examined risks of congenital defects other than neural tube defects that were previously hypothesized to be reduced by the use of multivitamin supplements. The present data confirm reductions in the risks of cleft palate alone, limb reduction defects, and urinary tract defects. Only moderate and statistically nonsignificant reductions in risk were observed for CL/P, hydrocephaly, and pyloric stenosis, and no reductions in risk were observed for conotruncal defects or ventricular septal defects.

Since multivitamins typically contain folic acid, multivitamin supplementation during the periconceptional period (pre-LM1) reflects behavior that is consistent with the US Public Health Service recommendation (1) that all women of childbearing age ingest 400 μg of folic acid daily to reduce the risk of neural tube defects. In the present study, such supplementation appeared to reduce the risks of limb reduction defects and urinary tract defects, and possibly the risks of CL/P, cleft palate alone, hydrocephaly, and pyloric stenosis as well, though estimates for the latter four defects were not statistically significant. In addition, supplementation that began after LM1 (after pregnancy was clinically recognizable) was associated with reduced risks of cleft palate alone and urinary tract defects. Furthermore, use that began in LM3 was associated with a reduction in pyloric stenosis risk that was of borderline statistical significance.

For neural tube defects, many studies have shown a multivitamin effect (1); however, it is widely held that it is the folic acid component that affords the benefit, because of a randomized trial conducted by the Medical Research Council (14) which showed a 60 percent reduction in risk of neural tube defects among women using folic acid supplements compared with women using multivitamins containing no folic acid. For defects other than neural tube defects, it is not clear what specific nutrient or combinations of nutrients might affect risk; in fact, it/they may vary from one malformation to another. Most vitamin supplements include more than eight water-soluble vitamins and three fat-soluble vitamins and at least four minerals or trace elements, and their overlap precluded us from identifying independent effects. In the present study, approximately 90 percent of nonprenatal multivitamin preparations and 100 percent of prenatal multivitamin preparations included folic acid.

Earlier reports on supplementation and reduced risks of malformations (other than neural tube defects) included a variety of study designs and definitions of supplementation (3–5, 15–18). The present observations support some but not all of those findings. For oral clefts, previous studies have reported inconsistent findings, including a reduced risk of CL/P (15); a reduced risk of cleft palate alone but not of CL/P (5, 16); a reduced risk of CL/P but not of cleft palate alone (17); reduced risks of both CL/P and cleft palate alone (3, 4); and no association for both CL/P and cleft palate alone (18). Differences in study design do not completely explain the inconsistencies in findings across the previous studies and the present data. One study conducted by our group using earlier data (18) showed no association, but we had included in the malformed control group some of the case defects found in the present study to be inversely associated with vitamin use. The inclusion of those defect groups as controls resulted in similar rates of supplementation between case and control groups, but this does not fully account for the earlier null finding.

In the present study, it appears that gestational timing may be an important factor for clefts: The greatest reduction in risk of CL/P (based on the malformed control group) was estimated for periconceptional use, whereas the greatest reduction in risk of cleft palate alone was found for first use in LM2 (using either control group). These findings are consistent with the fact that CL/P develops approximately 3 weeks earlier than cleft palate alone.

A randomized trial that treated approximately equal numbers of women with either multivitamins or trace elements showed the risk of ventricular septal defect to be lower in the vitamin-treated group (5). For conotruncal defects, risk was found to be reduced in two case-control studies and one randomized trial, but the effect was primarily observed for transposition of the great arteries (5, 6) and/or tetralogy of Fallot (5–7). In contrast, our data are consistent with those of a large case-control study (8) which showed no association between preconceptional multivitamin supplementation and risks of transposition of the great arteries and other conotruncal defects. Differences in study design offer no clear explanation for the discordant results. Furthermore, the distribution of defects within the conotruncal group in the present study was similar to that reported in the population-based studies (6–8): tetralogy of Fallot, 31 percent; transposition of the great arteries, 40 percent; truncus arteriosus, 11 percent; and double-outlet right ventricle, 10 percent.

For limb reduction defects, our findings are generally consistent with those of earlier reports which suggested reductions in risk for periconceptional or early first trimester use (5, 7, 10), with the greatest effect being evident for periconceptional use (5, 10). Two of the earlier studies divided limb reduction defects by type of limb deficiency; one found that the effect was confined to limb reduction defects other than longitudinal deficiencies (10), while the other found that the effect was confined to limb reduction defects other than transverse deficiencies (7). Unfortunately, there were too few limb reduction defects in the present study to examine these subtypes.

A case-control study of urinary tract defects and nonmalformed controls (9) found statistically significant reductions in risk associated with supplementation before, during, and after the first trimester. In addition, the previously described randomized trial (5) observed one case of obstructive urinary tract defects and no cases of renal agenesis in the supplemented group, as compared with three cases of the former and two cases of the latter in the nonsupplemented group. The present study confirmed these findings in that risks for urinary tract defects overall and for the subgroup of obstructive defects were generally below 1.0 for all timing categories and were significantly reduced for pre-LM1 and LM2 supplementation. There were too few cases of renal agenesis in the present study to examine them separately.

In the same randomized trial (5), fewer cases of congenital hydrocephalus and pyloric stenosis were observed in the supplemented group than in the nonsupplemented group. We estimated some risks of congenital hydrocephaly to be below the null value, but we found no statistically significant reductions in risk or informative patterns of risk across gestational timing categories. For pyloric stenosis, we found that periconceptional use or first use in LM2 did not reduce risk, but later first trimester use did. The possible benefit of relatively late exposure is consistent with the randomized trial in that women in the supplemented group were started on multivitamins before conception and use continued at least through the first trimester. Furthermore, the findings are of interest in that the gestational timing of pyloric stenosis is not known; but the defect is thought to result from a functional disorder of the pyloric sphincter, which suggests that it may arise later in gestation than most other congenital defects (19).

For the overall data, we considered possible sources of error. First, we attempted to reduce the potential for information bias with rigorous data collection, including use of a standardized questionnaire administered within 6 months after the date of delivery. Although the possibility of random misclassification of supplementation information still exists, we believe it would not have a strong influence on the findings, because our estimates of periconceptional prevalence for control mothers are similar to recently published rates (3, 20).

Differential misclassification (due either to maternal reporting or to interviewer bias) is another possibility, but it appears to be unlikely given that risk estimates based on nonmalformed controls were remarkably similar to those based on malformed controls. In addition, supplementation was reported to reduce the risks of many of the same defects in a randomized controlled trial (5) in which recall bias was not an issue.

There is a possibility that bias may have been introduced because subject ascertainment was not population-based, not all ascertained subjects were asked to participate, and approximately one third of mothers who were asked to participate refused to be interviewed. If incomplete enrollment were conditional on multivitamin use and such enrollment differed between cases and controls, selection bias would occur. However, the observed findings are not likely to be due to selection bias, for several reasons. First, the network of ascertainment hospitals included the full range of urban, suburban, and rural communities. Second, our use of general geographic proximity as the basis for choosing non-"priority" subjects had little effect on the distribution of socioeconomic status in each group, because the geographic areas were large and included communities representing all socioeconomic strata. Third, the similar distributions of zip code-linked median family income for US study subjects suggest that interviewed mothers may have had a higher socioeconomic status than noninterviewed mothers; but there was little difference in such status between cases and controls, which reduces concerns about biased risk estimation. Study procedures were the same at our Toronto center as at the other two centers, so the US zip code findings can most likely be extrapolated to our Canadian study subjects.

Selection bias might have been introduced by the inclusion of multivitamin-associated defect subgroups. The wide range of defects among the malformed controls and the similarity in supplementation rates between the two control groups suggest that selection bias was not introduced by the inclusion of these subgroups.

Finally, residual confounding is a possibility. We controlled for several factors that are associated with multivitamin supplementation, but there may be other differences between women who take supplements routinely or early in pregnancy and women who start taking them very late in pregnancy or do not take them

at all. If those differing factors affect risks of any of the case or control defects, observed estimates may have been confounded.

Studies carried out to date suggest that the benefits of periconceptional multivitamin supplementation may extend beyond a reduction in the risk of neural tube defects. If such supplementation does in fact reduce the risk of specific birth defects, the current public health recommendation that all women of childbearing age ingest 400 μg of folic acid daily (1) may need to be broadened to include a wider range of nutrients. Exactly which nutrients should be taken is not clear, but multivitamin supplements containing 400 μg of folic acid may offer greater benefit than folic acid supplements alone.

ACKNOWLEDGMENTS

Major support for this research was provided by grant RO1-HD27697 from the National Institute of Child Health and Human Development. Additional support for the Slone Epidemiology Unit Birth Defects Study was provided by Hoechst Marion Roussel, Inc. (Kansas City, Missouri), Pfizer, Inc. (New York, New York), and the Glaxo-Wellcome Company (Research Triangle Park, North Carolina).

The authors thank Irene Shepherd, interview coordinator; Rachel Wilson, hospital coordinator; Susan Littlefield and Sally Perkins, interviewers; Stephen Rivers, Joan Shander, and Diane Gallagher, research assistants; Thomas Kelley, research pharmacist; and Nastia Dynkin, programmer, for their assistance.

The authors are also indebted to the following institutions, which provided access to their patients: *Boston, Massachusetts*—Baystate Medical Center, Beth Israel Hospital, Boston City Hospital, Boston Regional Medical Center, Brigham and Women's Hospital, Brockton Hospital, Cambridge Hospital, Children's Hospital, Charlton Memorial Hospital, Deaconess Waltham Hospital, Emerson Hospital, Falmouth Hospital, Franciscan Children's Hospital, Good Samaritan Medical Center, Haverhill Municipal-Hale Hospital, Holy Family Hospital, Jordan Hospital, Kent County Memorial Hospital, Lawrence General Hospital, Lowell General Hospital, Malden Hospital, Medical Center of Central Massachusetts, Melrose-Wakefield Hospital, Metro West Medical Center (Framingham campus), Mt. Auburn Hospital, New England Medical Center, Newton-Wellesley Hospital, North Shore Medical Center, Rhode Island Hospital, Saints Memorial Medical Center, South Shore Hospital, Southern New Hampshire Regional Medical Center, St. Elizabeth's Medical Center, St. Luke's Hospital, St. Vincent Hospital, and Women and Infants' Hospital; *Philadelphia, Pennsylvania*—Abington Memorial Hospital, Albert Einstein Medical Center, Alfred I. du Pont Institute, Allegheny University Hospitals, Atlantic City Medical Center, Bryn Mawr Hospital, Chester County Hospital, Children's Hospital of Philadelphia, Community General Hospital, Cooper Hospital, Crozer-Chester Medical Center, Doylestown Hospital, Frankford Hospital (Torresdale Division), Hospital of the University of Pennsylvania, Lancaster General Hospital, Lehigh Valley Hospital, Medical Center of Delaware, Mercer Medical Center, Nanticoke Memorial Hospital, Our Lady of Lourdes Hospital, Pennsylvania Hospital, Rancocas Hospital, Reading Hospital and Medical Center, Sacred Heart Hospital, St. Mary Medical Center, Temple University Hospital, Thomas Jefferson University Hospital, and West Jersey Hospital; *Toronto, Ontario, Canada*—Cambridge Memorial Hospital, Guelph General Hospital, The Hospital for Sick Children, Joseph Brant Memorial Hospital, Kitchener-Waterloo Hospital, McMaster University Medical Center, Mississauga Hospital, Mt. Sinai Hospital, North York General Hospital, Oakville Trafalgar Memorial Hospital, Oshawa Hospital, Peel Memorial Hospital, Scarborough General Hospital, St. Joseph's Health Centre (London), St. Joseph's Health Centre (Toronto), St. Joseph's Hospital (Hamilton), St. Michael's Hospital, Toronto East General Hospital, Toronto General Hospital, Victoria Hospital, Women's College Hospital, York County Hospital, York Central Hospital, and York Finch Hospital.

REFERENCES

1. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neutral tube defects. MMWR Morb Mortal Wkly Rep 1992;41:1–7.
2. Oakley GP Jr, Erickson JD, Adams MJ Jr. Urgent need to increase folic acid consumption. (Editorial). JAMA 1995;274:1717–18.
3. Shaw GM, Lammer EJ, Wasserman CR, et al. Risks of orofacial clefts in children born to women using multivitamins containing folic acid periconceptionally. Lancet 1995;346:393–6.
4. Mills JL, McPartlin JM, Kirke PN, et al. Homocysteine metabolism in pregnancies complicated by neural-tube defects. Lancet 1995;345:149–51.
5. Czeizel AE. Prevention of congenital abnormalities by periconceptional multivitamin supplementation. BMJ 1993;306:1645–8.
6. Botto LD, Khoury MJ, Mulinare J, et al. Periconceptional multivitamin use and the occurrence of conotruncal heart defects: results from a population-based, case-control study. Pediatrics 1996;98:911–17.
7. Shaw GM, O'Malley CD, Wasserman CR, et al. Maternal periconceptional use of multivitamins and reduced risk for conotruncal heart defects and limb deficiencies among offspring. Am J Med Genet 1995;59:536–45.
8. Scanlon KS, Ferencz C, Loffredo CA, et al. Periconceptional folate intake and malformations of the cardiac outflow tract. Baltimore-Washington Infant Study Group. Epidemiology 1998;9:95–8.
9. Li DK, Daling JR, Mueller BA, et al. Periconceptional multivitamin use in relation to the risk of congenital urinary tract anomalies. Epidemiology 1995;6:212–18.
10. Yang Q, Khoury MJ, Olney RS, et al. Does periconceptional multivitamin use reduce the risk for limb deficiency in offspring? Epidemiology 1997;8:157–61.
11. Mitchell AA, Rosenberg L, Shapiro S, et al. Birth defects related to use of Bendectin in pregnancy. I. Oral clefts and cardiac defects. JAMA 1981;245:2311–14.

12. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
13. Bureau of the Census, US Department of Commerce. 1990 Census Lookup. (Website). http://www.venus.census.gov/cdrom/lookup/910817019.
14. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. Lancet 1991;338:131–7.
15. Tolarova M, Harris J. Reduced recurrence of orofacial clefts after periconceptional supplementation with high-dose folic acid and multivitamins. Teratology 1995;51:71–8.
16. Czeizel AE, Hirschberg J. Orofacial clefts in Hungary: epidemiological and genetic data, primary prevention. Folia Phoniatr Logop 1997;49:111–16.
17. Khoury MJ, Gomez-Farias M, Mulinare J. Does maternal cigarette smoking during pregnancy cause cleft lip and palate in offspring? Am J Dis Child 1989;143:333–7.
18. Hayes C, Werler MM, Willett WC, et al. Case-control study of periconceptional folic acid supplementation and oral clefts. Am J Epidemiol 1996;143:1229–34.
19. O'Rahilly R, Muller F. Human embryology and tetralogy. 2nd ed. New York, NY: John Wiley and Sons, Inc, 1996.
20. Knowledge and use of folic acid by women of childbearing age—United States, 1995. MMWR Morb Mortal Wkly Rep 1995;44:716–18.

J Clin Epidemiol Vol. 43, No. 1, pp. 87–91, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

RECALL BIAS IN EPIDEMIOLOGIC STUDIES

STEVEN S. COUGHLIN*

Department of Epidemiology, The Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland, U.S.A.

(*Received in revised form* 13 *April* 1989)

Abstract—The factors which contribute to bias due to differential recall between cases and controls in retrospective studies have been little studied. A review of the literature on recall accuracy suggests that the extent of inaccurate recall is related to characteristics of the exposure of interest and of the respondents, though a distinction must be drawn between recall which is biased and that which is simply inaccurate. Interviewing technique and the study protocol, including the design of questionnaires and the motivation of respondents, play a central role and are under the control of the investigator. The results of validation studies carried out to date suggest that the likelihood of recall bias may be greater when recall is poor in general.

Recall Bias Validity Misclassification Methods Study design

INTRODUCTION

It is generally accepted that comparative studies which attempt to retrospectively ascertain exposures through interviewing techniques may be subject to bias due to differential recall [1–3], and case–control studies which do not validate interview data are frequently criticized on this basis. However, the extent and nature of recall bias has been examined for only a small percentage of exposures and diseases, and relatively few methodologic reports of the factors which contribute to recall bias have been published.

Recall bias may be thought of as a form of differential misclassification bias and the risk estimate may be biased away from or towards the null [4–6]. Systematic error resulting from imperfect recall of exposure that is equally poor across case and comparison groups is perhaps more properly thought of as nondifferential misclassification bias. In this event, the risk estimate will generally be biased towards the null [4–6]. A distinction must be made between recall which is biased and that which is simply inaccurate [5]. Biased recall, which impacts on estimates of risk, implies a systematic departure from the truth. Thus, a consideration of recall bias in comparative studies should encompass both the extent to which recall of exposure is impaired and, more importantly, whether the impairment is different for cases and controls.

Although most of the published studies in this area address the accuracy of recall, their findings provide a useful foundation for a discussion of the factors which contribute to differential recall between cases and controls and systematic bias.

TIME INTERVAL AND DEGREE OF DETAIL

The time interval since exposure and the degree of detail required have been shown to influence the recall of a variety of exposures, including work histories [7, 8], dietary habits [9–13], obstetric histories [14–17], and medication usage [18–20]. For example, in a validation study of the accuracy of recall of spontaneous abortions, Wilcox and Horney found the recall of abortions occurring in the

*Reprint requests should be addressed to: Steven S. Coughlin, Ph.D., M.P.H., Division of Biostatistics and Epidemiology, Kober-Cogan Hall, Room 404, Georgetown University School of Medicine, Washington, DC 20007, U.S.A.

previous 10 years to be 82% complete, whereas the recall of abortions occurring 20 years or more prior to the interview was only 73% complete [14]. Similar findings were reported by Corwin *et al.* from their study of the ability of peptic ulcer patients to recall their past medical history [21]. The correlation between the accuracy of reported weight and time elapsed since its determination was 0.89 ($p < 0.01$).

The limited ability of individuals to recall detailed past events was illustrated by the validation study conducted by Bean *et al.* of the accuracy of reported menstrual histories [2, 16]. The percentage of women who correctly recalled within a year their age at menarche, age of natural or surgical menopause, and age at first use of oral contraceptives ranged from 75 to 90%. The percentage who correctly recalled their menstrual cycle length within one day, however, ranged from 0 to 60%.

PERSONAL CHARACTERISTICS

Characteristics of individuals such as age, education, and socioeconomic status have also been reported to influence recall accuracy [2, 19]. Stolley *et al.* found the percent agreement between patient-reported oral contraceptive histories and prescriber records to vary across groups defined by race, private vs public payment status, and age [20]. For instance, the percent agreement between reported and recorded name of latest product used was 91% for private patients and 77% for welfare or Medicaid patients ($p < 0.04$). The trends for age categories were not consistent. Yerushalmy *et al.* found that Hawaiian women over the age of 50 reported fewer abortions than younger age groups [17], although a cohort effect or temporal trend in the incidence of abortions in this population could also explain their findings [14].

In a validation study of work histories by Baumgarten *et al.*, the percent agreement was quite close across age, income, and education categories [7]. Stewart *et al.* found a U-shaped age effect in a validation study of self-reported work histories among shipyard workers [8]. The workers who were 65–69 years of age at the time of interview had the lowest adjusted percent agreement between reported and recorded year of hire, length of employment, and job title. For instance, the agreement between reported and recorded job title was 53% for workers who were 65–69 years of age, adjusted for race, age at first employment, years since leaving shipyard, length of employment, and year of hire, as compared to 60% for workers who were less than 65 years of age and 73% for workers who were 70 years of age or older.

Thus, a consistent relationship between the accuracy of recall and demographic factors has not been found, perhaps reflecting differences in the study populations, the questions being asked, or the nature of the exposure.

SIGNIFICANCE OF EVENTS

Several lines of evidence suggest that recall ability is also related to the significance of a past event. Survey scientists have observed that the frequency, duration, vividness and meaningfulness of an event contribute to recall [22, 23]. For instance, the psychologic and economic importance of oral contraceptives may contribute to the ability of users to recall product names and dates of use [18, 20]. Wilcox and Horney found that the major determinant of spontaneous abortion recall was the length of pregnancy at the time of abortion [14]. Only 54% of abortions which occurred within the first 6 weeks of gestation were recalled, compared to 93% of those occurring after 13 weeks. Recall of past occupations has been found to be positively associated with the duration of past employment [8], and negatively associated with the number of jobs held [7].

SOCIAL DESIRABILITY

The recall and reporting of exposures which are felt to be socially undesirable, such as cigarette smoking and drug usage, may be reduced [24–26]. In addition, the accuracy of self-reported events may be less among individuals who do not admit to socially undesirable behavior. Self-reported events tend to be distorted in a socially desirable direction and behaviors which are associated with a social stigma or are perceived to be personally threatening are often under-reported [24–28]. For example, cigarette smoking, especially among adolescents or announced quitters, may be denied or minimized [29–31]. Validation studies which have utilized biochemical markers of cigarette smoking, such as cotinine or thiocyanate, have documented appreciable misclassification of self-reported cigarette usage [29–31]. In a recent study of idiopathic dilated cardiomyopathy [32], the percent agreement between reported and recorded history of cigarette smoking was 71.7% among cases, as compared to 89.0, 87.2 and 90.2% for history of asthma, hypertension, and diabetes, respectively.

Paganini-Hill and Ross examined the validity of self-reported medication histories as part of a case–control study of breast cancer among women living in an affluent retirement community [19]. The correspondence between reported and recorded usage varied greatly across drug categories, with barbiturates having the lowest agreement. The view of barbiturates as an overused drug and the negative image associated with its use may have resulted in underreporting [19].

INTERVIEWING TECHNIQUE

Interviewing techniques have also been shown to influence the recall of past events [2, 26]. The content and form of questions may affect recall accuracy [2, 19, 33]. Supplementary devices such as introductions to sections of the questionnaire may increase responses, possibly because of the stimulus and time provided to the respondents [26]. In a validation study of recall of dental X-rays, Preston-Martin *et al.* utilized carefully devised probes and specific questions to assist recall [33]. The percent agreement between patient interviews and dental charts in their study was much greater than that reported in a previous study of diagnostic X-rays [33, 34]. An appropriate setting for the interview may also improve recall by providing stimuli which elicit the desired memory [19, 26, 33]. The motivation of the respondents to participate and events which occurred prior to the interview may also affect the quality of the responses [26, 33]. The accuracy and completeness of the information obtained may be improved by supplying motivation, such as in conveying the knowledge to the subjects that they may be making an important contribution to disease prevention or treatment [26].

Survey researchers have identified two types of memory errors that may occur: errors of omission, such as forgetting the use of a medication, and compression or telescoping of time, so that events are recalled as having occurred more recently than they actually did [26, 35]. Techniques developed to reduce such memory errors include aided recall, or providing the respondents with a list of possible responses and asking them to consult written records [21, 35, 36]. As an example, the subjects may be provided with lists of commercial product names or shown pictures of pharmaceutical tablets or other exposures during personal interviews [36]. Food models have been used in diet history surveys to determine the amounts of usual servings and to prompt further responses [12].

DIFFERENTIAL RECALL

The factors which contribute to differential recall between cases and controls have been less well studied. Some of the aforementioned factors which influence recall in general may be important. For instance, the motivation to participate may be greater for cases than for controls [5]. Furthermore, past exposures may be more vivid or meaningful to cases, possibly because of their awareness of potential risk factors for their condition or because of repeated interviewing by physicians. Conversely, controls may have had less contact with health care providers and be less sensitized to questions about such exposures as cigarette smoking.

Jain *et al.* examined the validity of a diet history questionnaire designed to assess the prediagnosis dietary intake of cancer patients and their controls [12]. The retrospective estimates of nutrient intake for colon cancer cases, after approximately 6 months, were found to be too low as a result of the subjects being influenced by their current diet. However, the estimates for the control subjects, who had not undergone changes in their diet, were found to be more accurate. Stolley *et al.* found the percent agreement between reported and recorded oral contraceptive histories to be different for cases of thromboembolism and their hospital controls [20]. The cases had a higher percent agreement compared to the controls for both the name of the product (92 vs 89%) and the total duration of use (47 vs 31%). The cases may have thought about their past oral contraceptive use on more occasions and to a greater extent. Klemetti and Saxen conducted a validation study of recall of environmental exposures during pregnancy by comparing retrospective and prospective interview data [37]. Unexpectedly, the agreement between the two exposure histories was comparable for mothers of normal children and for mothers of dead or malformed children.

Although differential errors of recall may inflate the risk estimate [5], the extent to which this occurs in case–control studies has not been examined. Figure 1 summarizes the results of four validation studies which examined the agreement between recalled and recorded exposures for both cases and controls [7, 15, 19, 20]. The overall percent agreement

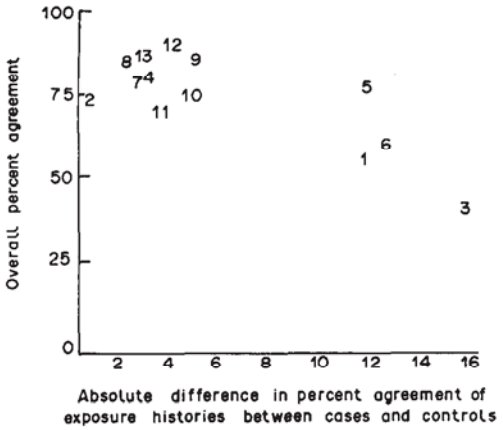

Fig. 1. Summary of results from validation studies of recall accuracy; overall percent agreement of reported and recorded exposure histories vs absolute difference in percent agreement between cases and controls. 1—Cases of thromboembolism and hospital controls, agreement between reported and recorded oral contraceptive starting date, Stolley *et al.* [20]; 2—agreement between reported and recorded oral contraceptive stopping date, Stolley *et al.* [20]; 3—agreement between reported and recorded duration of use of oral contraceptives, Stolley *et al.* [20]; 4—Cancer cases and controls, agreement between reported and recorded employer names, Baumgarten *et al.* [7]; 5—Breast cancer cases and community controls, agreement between reported and recorded ever/never use of oral estrogens, interview vs medical record, Paganini-Hill and Ross [19]; 6—agreement between reported and recorded ever/never use of oral estrogens, interview vs pharmacy records, Paganini-Hill and Ross [19]; 7—agreement between reported and recorded use of thyroid medication, Paganini-Hill and Ross [19]; 8—agreement between reported and recorded use of reserpine, Paganini-Hill and Ross [19]; 9—agreement between reported and recorded use of other antihypertensives, Paganini-Hill and Ross [19]; 10—agreement between reported and recorded use of steroids, Paganini-Hill and Ross [19]; 11—agreement between reported and recorded use of barbiturates, Paganini-Hill and Ross [19]; 12—DES-exposed mothers and DES-unexposed mothers, agreement between reported and recorded number of prior pregnancies, Tilley *et al.* [15]; 13—agreement between reported and recorded number of prior miscarriages, Tilley *et al.* [15].

between reported exposures obtained by interview and exposures recorded in the medical record, for cases and controls combined, is plotted against the absolute difference between cases and controls in the percent agreement between reported and recorded exposures. A variety of diseases and exposures are represented. Although the number of observations is somewhat small, there appears to be a roughly linear relationship between the overall percent agreement for cases and controls combined, with respect to reported and recorded exposures, and the absolute difference in percent agreement between cases and controls (Fig. 1). The largest differences in percent agreement between cases and controls were seen in those instances where the overall percent agreement was relatively poor. Thus, systematic errors in recall of exposures may be more likely to occur in comparative studies when inaccuracies of recall are more frequent in general.

Approaches to study design that may minimize differential recall of exposures include the use of controls who are likely to have considered past exposures to the same extent as the cases [36, 38]. For example, the use of cancer controls in studies of specific cancer sites may increase the comparability of the information obtained from cases and controls, and enable the investigators to control for recall bias [38]. The aforementioned interviewing techniques which improve recall in general, such as the provision of memory aids, may also lessen the possibility of recall bias [19, 26, 33, 35, 36].

SUMMARY AND CONCLUSIONS

In summary, the extent of inaccurate recall in retrospective studies is determined by characteristics of the exposure of interest including the degree of detail, significance to the respondent, social acceptance and time period involved. Interviewing technique and the study protocol, including the design of questionnaires and the motivation of respondents, play a central role and are under the control of the investigator. Respondent characteristics also contribute to the extent of error, but this influence may be less predictable. The contribution of these factors to the extent of differential recall between cases and controls has been examined infrequently, and more research in this area is needed.

The results of validation studies carried out to date suggest that the likelihood of recall bias may be greater when recall is generally poor (Fig. 1). Nonetheless, several well designed studies have looked for and failed to find appreciable evidence of inaccurate recall [7, 33] or differential recall [37, 39]. Future studies should address those aspects of study design and questionnaire development which improve recall and, in particular, lessen the potential for differential recall between cases and controls.

REFERENCES

1. Lilienfeld AM, Lilienfeld D. **Foundations of Epidemiology**, 2nd Edn. New York: Oxford University Press; 1980.
2. Schlesselman JJ. **Case–Control Studies. Design, Conduct, Analysis.** New York: Oxford University Press; 1982.
3. Sackett DL. Bias in analytic research. **J Chron Dis** 1979; 32: 51–63.

4. Copeland KT, Checkoway H, McMichael AJ, Holdbrook RH. Bias due to misclassification in the estimation of relative risk. **Am J Epidemiol** 1977; 105: 488–495.
5. Raphael K. Recall bias: A proposal for assessment and control. **Int J Epidemiol** 1987; 16: 167–170.
6. Barron BA. Effects of misclassification on the estimation of relative risk. **Biometrics** 1977; 33: 414–418.
7. Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. **Am J Epidemiol** 1983; 118: 583–591.
8. Stewart W, Tonascia J, Matanoski G. The validity of questionnaire-reported work history in live respondents. **J Occup Med** 1987; 29: 795–800.
9. Rohan TE, Potter JD. Retrospective assessment of dietary intake. **Am J Epidemiol** 1984; 120: 876–887.
10. Van Leeuwen FE, De Vat HCW, Hayes RB *et al.* An assessment of the relative validity of retrospective interviewing for measuring dietary intake. **Am J Epidemiol** 1983; 118: 752–758.
11. Jensen OM, Wahrendorf J, Rosenqvist A, Geser A. The reliability of questionnaire-derived historical dietary information and temporal stability of food habits in individuals. **Am J Epidemiol** 1984; 120: 281–290.
12. Jain M, Howe GR, Johnson KC, Miller AB. Evaluation of a diet history questionnaire for epidemiologic studies. **Am J Epidemiol** 1980; 111: 212–219.
13. Vobecky JS, Vobecky J, Froda S. The reliability of the maternal memory in a retrospective assessment of nutritional status. **J Clin Epidemiol** 1988; 41: 261–265.
14. Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. **Am J Epidemiol** 1984; 120: 727–733.
15. Tilley BC, Barnes AB, Bergstralh E *et al.* A comparison of pregnancy history recall and medical records. Implications for retrospective studies. **Am J Epidemiol** 1985; 121: 269–281.
16. Bean JA, Leeper JD, Wallace RB *et al.* Variations in the reporting of menstrual histories. **Am J Epidemiol** 1979; 109: 181–185.
17. Yerushalmy J, Bierman JM, Kemp DH *et al.* Longitudinal studies of pregnancy on the island of Kaui, Territory of Hawaii. **Am J Obstet Gynecol** 1956; 71: 80–96.
18. Glass R, Johnson B, Vessey M. Accuracy of recall of oral contraceptive use. **Br J Prev Soc Med** 1974; 28: 273–275.
19. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. **Am J Epidemiol** 1982; 116: 114–122.
20. Stolley PD. Tonascia JA, Sartwell PE *et al.* Agreement rates between oral contraceptive users and prescribers in relation to drug use history. **Am J Epidemiol** 1978; 107: 226–235.
21. Corwin RG, Krober M, Roth HP. Patients' accuracy in reporting their past medical history. A study of 90 patients with peptic ulcer. **J Chron Dis** 1971; 23: 875–879.
22. Parten M. **Surveys, Polls, and Samples: Practical Procedures**. New York: Harper and Brothers; 1950.
23. Chappell MN. Factors influencing recall of radio programs. **Public Opinion Q** 1942; 6: 197–214.
24. Warner KE. Possible increases in the under-reporting of cigarette consumption. **J Am Stat Assoc** 1978; 73: 314–316.
25. Bauman KE, Koch GG. Validity of self-reports and descriptive and analytical conclusions. The case of cigarette smoking by adolescents and their mothers. **Am J Epidemiol** 1983; 118: 90–98.
26. Rossi PH, Wright JD, Anderson AB, Eds. **Handbook of Survey Research**. New York: Academic Press; 1983.
27. Parry HH, Crossley HM. Validity of responses to survey questions. **Public Opinion Q** 1950; 14: 61–80.
28. Backstrom CH, Hursh-Ceaser G. **Survey Research**, 2nd edn. New York: John Wiley; 1981.
29. Haley NJ, Axelrod CM, Tilton KA. Validation of self-reported smoking behavior: biochemical analyses of continine and thiocyanate. **Am J Public Health** 1983; 73: 1204–1207.
30. Vasey CJ, Saboojee Y, Cole PV, Russell MAH. Blood carboxyhemoglobin, plasma thiocyanate and cigarette consumption: implications for epidemiological studies in smokers. **Br Med J** 1982; 284: 1516–1518.
31. Gillies PA, Wilcox B, Coates C *et al.* Use of objective measurement in the validaton of self-reported smoking in children aged 10–11 years: saliva thiocyanate. **J Epidemiol Commun Health** 1982; 36: 205–208.
32. Coughlin SS. A Case–Control Study of Dilated Cardiomyopathy. Doctoral dissertation, The Johns Hopkins University; 1987.
33. Preston-Martin S, Bernstein L, Maldonado AA *et al.* A dental X-ray validation study. **Am J Epidemiol** 1985; 121: 430–439.
34. Graham S, Levin ML, Lilienfeld AM *et al.* Methodological problems and design of the tristate leukemia study. **Ann NY Acad Sci** 1963; 107: 557–569.
35. Sudman S, Bradburn N. **Response Effects in Surveys: A Review and Synthesis**. Chicago: Aldine; 1974.
36. Feinstein AR. **Clinical Epidemiology. The Architecture of Clinical Research.** Philadelphia: W.B. Saunders Co., 1985.
37. Klemetti A, Saxen L. Prospective versus longitudinal approach in the search for environmental causes of malformation. **Am J Public Health** 1967; 57: 2071–2075.
38. Linet MS, Brookmeyer R. Use of cancer controls in case–control cancer studies. **Am J Epidemiol** 1987; 125: 1–11.
39. Mackenzie SG, Lippman A. An investigation of report bias in a case–control study of pregnancy outcome. **Am J Epidemiol** 1989; 129: 65–75.