



Can Men Be Trusted? A Comparison of Pregnancy Histories Reported by Husbands and Wives

Fariyal F. Fikree,¹ Ronald H. Gray,² and Farida Shah²

Agreement between pregnancies and pregnancy outcomes reported by husbands and wives was assessed in a sample of 857 couples interviewed between June 1989 and July 1990. The respondents were men employed in a semiconductor manufacturing plant in Burlington, Vermont, and their wives. The wives' reports were used as the standard against which the husbands' reproductive histories were evaluated. Measures included sensitivity, specificity, and percentage of agreement. Reports were considered to be congruent if an outcome reported by the husband agreed with the outcome reported by the wife within a period of ± 6 months. Although men and women reported similar numbers of livebirths (1,478 and 1,500, respectively), men tended to misreport the timing of events; therefore, complete agreement on the numbers and dates of births was only 88.5%. Men also misreported the prevalence of low birth weight (sensitivity, 74%). Specificity was poorer for the younger (<35 years) and less educated (≤ 12 years) respondents. Husbands' reports of spontaneous abortions had lower sensitivity (71.2%) than their reports of livebirths, particularly among the better educated (66.9%). Induced abortions were frequently omitted by the husbands (sensitivity, 35.1%), and events such as stillbirths or tubal pregnancies were too few in number to permit meaningful analysis. It is concluded that husbands' misreporting of their wives' reproductive histories may be substantial and sufficient to compromise the validity of epidemiologic studies. It would, therefore, be prudent to avoid the use of husbands as proxy informants of their wives' reproductive histories. *Am J Epidemiol* 1993;138:237-42.

epidemiologic methods; medical history taking; pregnancy outcome; sensitivity and specificity (epidemiology)

Epidemiologic investigations often use spouses as surrogates to provide proxy information when the index partner is not available or is thought to be an unreliable respondent. Studies that utilize surrogate information include those in which the partner is deceased (1, 2), unavailable (3, 4), or un-

reliable (e.g., because of neurologic, psychiatric, or pediatric conditions) (5, 6). Nelson et al. (7) recently reviewed the use of proxy information on exposure, but there has been little attention given to the use of proxy information on outcome variables, particularly those related to reproduction.

The reliability of men as proxy respondents for provision of information on reproductive health, especially reports of spontaneous abortions, is particularly important for research on occupational hazards to which men are exposed (8, 9) and thus may be more readily available for interview. However, reports by men may be unreliable because of failure to recall pregnancies and pregnancy outcomes, inability to accurately locate these events in time,

Received for publication September 4, 1992, and in final form May 14, 1993.

¹ Department of Community Health Sciences, The Aga Khan University, Stadium Road, P. O. Box 3500, Karachi 74800, Pakistan.

² Department of Population Dynamics, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD.

Reprint requests to Dr. Ronald H. Gray, Department of Population Dynamics, The Johns Hopkins University School of Hygiene and Public Health, 615 N. Wolfe Street, Baltimore, MD 21205.

and a lack of knowledge of events (e.g., abortion) or characteristics of outcomes, such as birth weight and gestational age. The purpose of this report is to compare reproductive histories given by husbands and wives in order to assess whether interviewing men alone would provide reliable estimates of pregnancy outcomes.

This paper evaluates the completeness of information given by husbands as compared with that of their wives on the occurrence of livebirths, spontaneous abortions, induced abortions, and birth weights provided by 857 partner pairs in a study of semiconductor manufacturing employees.

MATERIALS AND METHODS

The study of reproductive outcomes was conducted between June 1989 and July 1990 in a semiconductor manufacturing plant in Burlington, Vermont. Full reproductive histories were obtained independently from 20- to 45-year-old men who were employees of the plant and from their wives, irrespective of whether the latter were employed at the plant. Interviews were conducted in person by trained interviewers using a computer-assisted questionnaire.

A total of 1,144 men agreed to be interviewed. However, 150 wives (13.1 percent) refused interview, and 137 couples (12.0 percent) were not currently married at the time of interview, leaving a sample of 857 currently married couples for whom there were completed interviews for both partners. The reproductive histories included information on the dates, number, and outcome of pregnancies and on birth weight for livebirths. In addition, the beginning and ending dates of previous marriages or relationships for each partner were obtained. Husbands and wives provided information independently on 2,055 and 2,197 pregnancies, respectively. Restricting the data to pregnancy outcomes during the respondents' current marriage resulted in 1,833 pregnancy outcomes reported by wives and 1,728 reported by their husbands.

To assess agreement between husbands

and wives, we made the following assumptions. First, the outcomes reported by husbands and wives were considered to be in agreement if the pregnancy order corresponded and if the date of the outcome reported by the couple agreed within ± 6 months. Otherwise the reported outcomes were treated as discrepant.

Second, if either or both partners reported a previous relationship, then the date of termination of the previous relationship had to precede the date of conception for the pregnancy of interest. Otherwise it was excluded, as there was no way to determine whether the pregnancy belonged to the previous or current relationship.

The events reported by women were assumed to be accurate and were used as the standard by which reports from men were assessed. (We traced medical records for 82.1 percent of miscarriages reported by the women and confirmed the maternal report in 97.4 percent of cases.) Statistical measures used to assess the validity of husbands' reports included sensitivity, specificity, and percentage of agreement (10). Sensitivity was estimated from the index events reported by both husbands and wives (true positives detected) relative to the total index events reported by the wives. Specificity was estimated from non-index events reported by both the husbands and wives relative to all non-index events reported by the wives. Statistical tests, such as the χ^2 and Z tests, were used to assess the statistical significance of differences in proportional distributions (11). Kappa statistics were not used, because the woman's report was taken as the standard. Data were stratified by time between pregnancy and interview and by the age and education of respondents.

RESULTS

Fifty-five percent of husbands and 44.7 percent of their wives were aged 35 years or older, and the majority had more than a high school education (men, 56.8 percent; women, 51.8 percent). Almost all the respondents were white (98.5 percent), and

approximately 80 percent had only one relationship. Table 1 shows the reproductive histories reported by men and women. Although the proportional distributions of pregnancy outcomes were similar ($\chi^2 = 1.1$), men reported fewer events than did women, particularly spontaneous and induced abortions. Men, however, over-reported the number of low-birth-weight infants, but this was not statistically significant.

Despite similarities in the number of livebirths reported by husbands ($n = 1,478$) and wives ($n = 1,500$), there were discrepancies in the timing of reported events. Men tended to misreport the year of birth. There were 1,466 livebirths in which both spouses reported a date of birth. Complete agreement was observed in 88.5 percent of cases, whereas agreement within ± 3 months was 91.6 percent, agreement within ± 6 months was 92.2 percent, and agreement within ± 12 months was 98.8 percent. Agreement on the timing of events was also poor for other pregnancy outcomes. For example, with spontaneous abortions, complete agreement on the month and year of the event was only 31.7 percent, and agreement within ± 6 months was only 65.5 percent.

When livebirths were matched within a

± 6 -month time period, there was moderate agreement between reports of livebirths by men and women, irrespective of the respondent's age and the time since the event (table 2). However, specificity was much lower for respondents with a high school education (65.5 percent) compared with those with tertiary education (81.6 percent), and this difference was highly significant ($p \leq 0.001$). The information on low birth weight obtained from men is also shown in table 2. Sensitivity was significantly lower for younger and less educated respondents.

For spontaneous abortions, sensitivity was approximately 70 percent, and specificity was approximately 99 percent irrespective of the time since outcome (table 3). There were no overall differences by age of respondent, but sensitivity was lower for more educated respondents as compared with those with 12 or less years of education (66.9 percent and 77.8 percent, respectively; $p = 0.06$).

There was poor agreement between reports from men and women of induced abortion, and sensitivity was only 35.1 percent (table 3). Husbands' reports of stillbirths and tubal pregnancies showed substantial agreement, but analysis was limited by the small numbers (results not shown).

TABLE 1. Pregnancy outcomes reported by husbands and wives: Burlington, Vermont, June 1989 to July 1990*

Reported outcome	Events reported by male interviewees ($n = 1,728$)		Events reported by female interviewees ($n = 1,833$)	
	No.	%	No.	%
Livebirths	1,478	85.5	1,500	81.8
Stillbirths	10	0.6	9	0.5
Spontaneous abortions	198	11.5	250	13.6
Induced abortions	32	1.8	57	3.1
Tubal pregnancy	10	0.6	15	0.8
Birth weight (g)†				
<2,500	84	5.7	73	4.9
$\geq 2,500$	1,394	94.3	1,427	95.1

* The total number of couples interviewed was 857.

† Based on the reported numbers of livebirths.

DISCUSSION

The objective of this study was to determine whether reproductive histories obtained from employed, married men could be used as substitutes for histories obtained from women in research on occupational hazards. It was assumed that the histories provided by women represented the most reliable information, because previous studies have shown that women recall with acceptable accuracy the number of births and pregnancy outcomes, such as spontaneous abortions or low birth weight (12, 13).

We could not find other published studies comparing husbands' reports with their wives' pregnancy histories, although several investigators have examined spousal proxy

TABLE 2. Sensitivity and specificity of husbands' reports of livebirths and low birth weight by time since event, age, and education: Burlington, Vermont, June 1989 to July 1990

Variable	No. of index events reported by wives*	No. of index events reported by husbands*	No. of non-index events reported by wives*	No. of non-index events reported by husbands*	Sensitivity† (%)	Specificity‡ (%)
Livebirths						
Total no.	1,500	1,351	486	361	90.1	74.3
Time since event (years)						
0-5	452	410	149	110	90.7	73.8
6-10	490	449	168	129	91.6	76.8
>10	558	492	169	122	88.2	72.2
Age (years)						
<35	798	720	266	192	90.2	72.2
≥35	702	631	220	169	89.9	76.8
Education (years)						
≤12	747	654	220	144	87.6	65.5
>12	753	697	266	217	92.6	81.6
Low birth weight						
Total no.	73	54	1,552	1,528	74.0	98.5
Time since event (years)						
0-5	13	10	478	471	76.9	98.5
6-10	26	20	503	495	76.9	98.4
>10	34	24	571	562	70.6	98.4
Age (years)						
<35	36	24	836	823	66.7	98.4
≥35	37	30	716	705	81.1	98.5
Education (years)						
≤12	41	27	782	768	65.9	98.2
>12	32	27	770	760	84.4	98.7

* Index events were livebirths or low birth weight births, respectively, which were reported by the wives. All other pregnancy outcomes were non-index events.

† Sensitivity is the index events correctly reported by the husbands as a percentage of total index events reported by the wives.

‡ Specificity is the non-index events correctly reported by the husbands as a percentage of total non-index events reported by the wives.

information on exposures. Thorogood and Vessey (4) found husbands' reports of their wives' height, weight, smoking, and oral contraceptive use to be very reliable. Similarly, other investigators found that surrogate information on major illnesses, smoking, and recent occupation is generally reliable, whereas information on other exposures such as diet, alcohol, amount smoked, or more distant jobs was less satisfactory (1-3, 7).

Our findings suggest that reliance on pregnancy histories derived from interviews may be problematic for epidemiologic research on male exposures and reproductive outcomes. In this study, men reported the number of livebirths with reasonable accu-

racy, but they misreported the month and year of birth for liveborn children. Timing errors were also observed with spontaneous abortion. The timing of exposure prior to or during pregnancy may be related to pregnancy outcome, and exposures may vary over time; thus, the discrepancy in reported dates of births or miscarriage could potentially result in substantial misclassification. Such misclassification with respect to timing is likely to be nondifferential, and the resulting bias should be toward the null. Similarly, the lower sensitivity of husbands' reports of low birth weight and spontaneous and induced abortions could result in substantial omission of events. In summary, the errors by men in reporting reproductive

TABLE 3. Sensitivity and specificity of husbands' reports of spontaneous and induced abortions by time since event, age, and education: Burlington, Vermont, June 1989 to July 1990

Variable	No. of index events reported by wives*	No. of index events reported by husbands*	No. of non-index events reported by wives*	No. of non-index events reported by husbands*	Sensitivity† (%)	Specificity‡ (%)
Spontaneous abortion						
Total no.	250	178	1,627	1,607	71.2	98.8
Time since event (years)						
0-5	89	61	475	466	68.5	98.1
6-10	79	56	546	540	70.9	98.9
>10	82	61	606	601	74.4	99.2
Age (years)						
<35	125	89	878	868	71.2	98.9
≥35	125	89	749	739	71.2	98.7
Education (years)						
≤12	99	77	802	790	77.8	98.5
>12	151	101	825	817	66.9	99.0
Induced abortion						
Total no.	57	20	1,820	1,808	35.1	99.3
Time since event (years)						
0-5	7	2	557	555	28.6	99.6
6-10	26	10	599	594	38.5	99.2
>10	24	8	664	659	33.3	99.2
Age (years)						
<35	43	16	960	952	37.2	99.2
≥35	14	4	860	856	28.6	99.5
Education (years)						
≤12	20	8	881	878	40.0	99.7
>12	37	12	939	930	32.4	99.0

* Index events were livebirths or low birth weight births, respectively, which were reported by the wives. All other pregnancy outcomes were non-index events.

† Sensitivity is the index events correctly reported by the husbands as a percentage of total index events reported by the wives

‡ Specificity is the non-index events correctly reported by the husbands as a percentage of total non-index events reported by the wives

events appear to vary by type of pregnancy outcome and the characteristics of respondents, and such errors could affect the results of epidemiologic studies.

Brenner and Savitz (14) have shown that a high level of specificity for outcome measures is crucial to the validity of case-control studies. Lower specificity can compromise power and, in general, the odds ratios are biased toward the null. Poor sensitivity is less of a problem in this context, although it may compromise precision (14). Also, Flegal et al. (15) have shown that imperfect sensitivity or specificity can underestimate associations in cross-sectional or cohort studies, leading to a biased underestimate of relative risk. Similarly, Kupper (16) found

that unreliability of surrogate confounders can seriously affect the validity of results.

In summary, our analysis suggests that errors in the reporting of reproductive events by men may be substantial and sufficient to compromise study validity. Therefore, where possible, it would be prudent to avoid the use of husbands as proxy informants of their wives' reproductive histories.

ACKNOWLEDGMENTS

Support for this study was provided by grant IBM H32.7936.12.

The authors thank William Hou for computer programming assistance and Dr. Rosemarie

Hakim and Dr. Penelope Keyl for their comments on the manuscript.

REFERENCES

1. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 1986;123:481-9.
2. Pickle LW, Brown LM, Blot WJ. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol* 1983;118:99-108.
3. Hatch MC, Misra D, Kabat GC, et al. Proxy respondents in reproductive research: a comparison of self- and partner-reported data. *Am J Epidemiol* 1991;133:826-31.
4. Thorogood M, Vessey M. The reliability of surrogate information about oral contraceptive use, smoking, height, and weight collected from men about their wives. *Contraception* 1989;39:401-8.
5. Rocca WA, Fratiglioni L, Bracco L, et al. The use of surrogate respondents to obtain questionnaire data in case-control studies of neurologic diseases. *J Chronic Dis* 1986;39:907-12.
6. Heyman A, Wilkinson WE, Stafford JA, et al. Alzheimer's disease: a study of epidemiologic aspects. *Ann Neurol* 1984;15:335-41.
7. Nelson LM, Longstreth WT Jr, Koepsell TD, et al. Proxy respondents in epidemiologic research. *Epidemiol Rev* 1990;12:71-86.
8. Buffler PA, Aase JM. Genetic risks and environmental surveillance: epidemiological aspects of monitoring industrial populations for environmental mutagens. *J Occup Med* 1982;24:305-14.
9. Wilcox AJ. Surveillance of pregnancy loss in human populations. *Am J Ind Med* 1983;4:285-91.
10. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley & Sons, Inc, 1981:212-25.
11. Kuzma JW. Basic statistics for the health sciences. Palo Alto, CA: Mayfield Publishing Company, 1984:135-7, 141-57.
12. Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol* 1984;120:727-33.
13. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol* 1989;129:233-48.
14. Brenner H, Savitz DA. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. *Am J Epidemiol* 1990;132:181-92.
15. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol* 1986;123:736-51.
16. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *Am J Epidemiol* 1984;120:643-8.