Misclassification Bias Related to Definition of Menopausal Status in Case-Control Studies of Breast Cancer

ALFREDO MORABIA AND PHILIPPE FLANDRE

Morabia A (Hôpital Cantonal Universitaire de Genève, Unité d'Épidémiologie Clinique, Rue Micheli-du-Crest 25, CH-1211 Geneva 4, Switzerland) and Flandre P. Misclassification bias related to definition of menopausal status in case-control studies of breast cancer. *International Journal of Epidemiology* 1992; **21**: 222–228.
It is often assumed, but has not been consistently observed, that some characteristics of reproductive history are specifically related to breast cancer of pre- or postmenopausal onset. To determine whether inconsistent reports may be due to differences in definition of menopause, we computed the relative odds (RO) of breast cancer for nulliparity, age at first live birth, family history of breast cancer and prior history of benign breast disease, separately in pre- and postmenopausal women, using seven different definitions of menopause. Results show that (i) relative odds of breast cancer and their confidence intervals may vary according to definitions of menopause; (ii) age-based definitions of menopause are associated with moderate differential misclassification bias between cases and controls; (iii) nulliparity, late age at first birth and family history of breast cancer seem to be specific risk factors for pre- but not postmenopausal breast cancer when cutoff for menopausal status is 10 years or more after last menses; and (iv) when information on menstrual history is not available, 50 years of age may be the best proxy for all menses-based definitions of menopause. We conclude that inconsistent findings on the effect of menopausal status in the association of breast cancer with some reproductive factors are partly due to statistical imprecision and differential misclassification bias associated with different age-based or menses-based definitions of menopause. Researchers should either test whether their conclusions hold using several definitions of menopause or give a biological rationale for the choice of a given definition of menopause.

It is reasonable to believe that hormonal changes related to menopause can play a role in the aetiology of breast cancer. Oestrogen and progestins are responsible for the proliferation of the mammary gland epithelium and for its differentiation.[1] Artificial or natural menopause may therefore affect the metabolism and susceptibility to carcinogens of the mammary gland since, at menopause, progestin secretion stops and oestrogen secretion occurs in the adrenals and in the adipose tissue.[2]

It has been suggested that premenopausal breast cancer may be primarily related to excess exposure of the mammary gland to ovarian secretion of sex hormones while environmental risk factors would be prime in the aetiology of postmenopausal breast cancer.[3-5] Since reproductive history determines a woman's life exposure to oestrogen and progestins, this theory implies that reproductive histories with longer exposures to ovarian hormones should increase

risk of breast cancer of premenopausal onset but not necessarily of breast cancer of postmenopausal onset.

However, epidemiological studies of breast cancer have not consistently shown that premenopausal and postmenopausal breast cancer were related to a different set of reproductive risk factors.[6] Inconsistencies across studies may stem from the considerable differences in definitions of menopause. Some authors have defined menopause according to time elapsed between last menses and diagnosis of breast cancer (menses-based definition).[7-19] Others, usually lacking information on menstrual history, used age as a proxy for menopausal status (age-based definition).[20-24] Finally, some, have used mixed definitions.[25-27] In addition, there is extensive variation within these two categories of definitions. The cutoff for menses-based definition of menopause varies from 0 (age at self-reported menses cessation) to 10 years after last menses, whereas ages between 45 and 55 years have been used as age-based definitions.

In the present study we therefore determined whether associations of breast cancer with postulated

Hôpital Cantonal Universitaire de Genève, Unité d'Épidémiologie Clinique, Rue Micheli-du-Crest 25, CH-1211 Geneva 4, Switzerland.

risk factors such as nulliparity, age at first live birth, family history of breast cancer and personal history of benign breast disease differ in premenopausal and postmenopausal women, using seven different definitions of menopause.

MATERIALS AND METHODS

Between 1973 and 1975, a case-control study was performed in Baltimore City and County and in the St Paul-Minneapolis area to determine the role of reproductive and endocrinological factors in the aetiology of breast cancer. The study has been described in detail elsewhere.[28] Briefly, cases were women histologically diagnosed as having primary malignant neoplasm of the mammary gland and free of prior breast cancer. Diagnoses of breast cancer were abstracted from pathology reports in five different Maryland and 12 different Minnesota area hospitals.

Hospital controls were women free of prior breast cancer, of the same age group (±5 years) and race, and admitted to the same hospital, during the same period (±2 months) as the case. Patients admitted to Obstetric-Gynaecology and Psychiatry services and those admitted for breast diseases were not eligible.

The present analyses are based on 427 cases and 560 hospital controls, after exclusion of women for whom date of last menses was missing or whose pregnancy outcomes had only been miscarriages, abortions or stillbirths. Women were classified as nulliparous if they never had a full-term pregnancy. Parous women were categorized according to their age at first live birth (early '<25 years', late '≥25 years'). A family history was considered positive when the mother or at least one sister had had breast cancer. Otherwise, family history was regarded as negative or unknown. History of benign breast disease was present when the subject reported having had a nonmalignant breast lump diagnosed by a physician.

Using menses-based definitions, a woman was categorized as postmenopausal if her last menses occurred >3 months, >1 year, >2 years or >10 years before diagnosis or interview. Five women who said they were premenopausal but had not had menses for 3–6 months because of menses irregularity, were classified as premenopausal. For age-based definitions of menopausal status, we used the cutoffs >45 years, >50 years or >55 years. We did not separate women with natural menopause from women with surgical menopause because the purpose of the study was to compare age-based with menses-based definitions of menopause and data on type of menopause are usually not available in the studies with age-based definitions of menopause.

Sensitivity and specificity of the three age-based definitions of menopause were computed relative to each menses-based definitions of menopause. For example, taking 'more than 1 year after last menses' as a menses-based definition of menopause, we computed the sensitivity and specificity of the 45, 50 and 55 years age-based definitions of menopause separately for cases and controls.

Distributions of ages at menopauses were smoothed using the 'ksmooth' procedure of S-plus.[29] Multivariate analyses were performed using unconditional logistic regression. Results are reported after adjustment for age (5-year age groups), state (Minnesota, Maryland), race (black, white) and education (≤12, 13-15, ≥16 years).

Analyses were repeated for the seven different definitions of menopause. Each model included an interaction term between the variable 'menopause' and the studied risk factor. The ratio of the coefficient of the interaction term over its standard error (SE) was used as test for interaction and considered as statistically significant at the 5% level if ≥1.96.[30]

RESULTS

Some 237 cases (55.5%) and 342 controls (61.1%) reported that their menses had ceased. Mean age at last menses was 45.5 (SD = 7.0) years in cases and 42.2 (SD = 8.5) years in controls (P<0.01). Figure 1 shows that the distribution of age at menopause was shifted to the right for cases compared to controls.
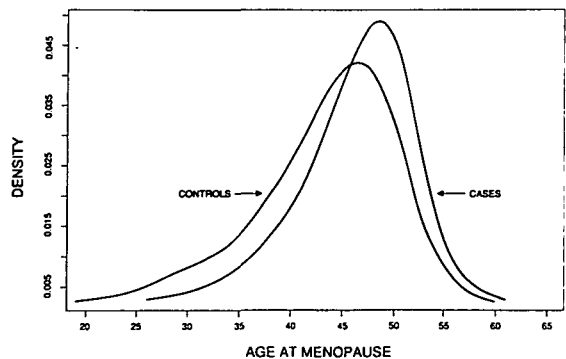

FIGURE 1   Density distribution of age at menopause in breast cancer cases and hospital controls. Maryland and Minnesota, 1973–1975.

Tables 1 and 2 present seven different stratifications of the total sample. Premenopausal women are those that do not meet a given definition of menopause. Table 1 shows sample sizes by risk factor for each of the seven definitions of menopause.

TABLE 1 *Distribution of breast cancer cases and hospital controls according to studied factors, by menopausal status and definition of menopause, Maryland and Minnesota, 1973-1975.*

Menopausal status	Whole sample		Nulliparity		Age at first live birth ≥ 25 years		Breast cancer in family		Prior breast lump	
Definition	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
	No.	No.	No.	No.	No.	No.	No.	No.	No.	No.
Premenopausal										
Months after last menses										
≤ 3	190	217	29	25	93	63	22	13	47	47
≤ 12	195	221	31	25	97	65	24	14	50	49
≤ 24	208	238	32	25	100	66	25	15	52	52
≤ 120	284	374	38	32	138	112	35	23	72	76
Age (years)										
≤ 45	175	260	24	26	80	67	21	13	41	58
≤ 50	220	327	30	28	97	86	28	16	55	77
≤ 55	270	383	38	36	129	110	33	20	70	83
Postmenopausal										
Months after last menses										
> 3	237	343	50	54	121	129	31	22	46	60
> 12	232	339	48	54	117	127	29	21	43	58
> 24	219	321	47	54	114	126	28	20	41	55
> 120	143	186	41	47	76	80	18	12	21	31
Age (years)										
> 45	252	300	55	53	134	125	32	22	52	49
> 50	207	233	49	51	117	106	25	19	38	30
> 55	157	177	41	43	85	82	20	15	23	24

In Table 2, relative odds (RO) for nulliparity vary little according to the definition used: between 1.4 and 1.7 in premenopausal women, and between 1.0 and 1.2 in postmenopausal women. Although none of the RO are statistically significant at the 5% level, the lower bound of the 95% confidence interval of the RO is 1.0 in premenopausal women within 120 months of their last menses or below age 56 years. For age at first live birth, RO in premenopausal women vary between 1.8 and 2.0 and are all statistically significant. In postmenopausal women, RO vary between 1.3 and 1.5 and are of borderline statistical significance. Relative odds for family history of breast cancer vary between 1.7 and 1.9 in pre- and postmenopausal women for menses-based definitions of menopause. In contrast, using the age-based definition, family history appears to be a risk factor in premenopausal women (RO between 2.3

and 2.7, statistically significant) but not in postmenopausal women (RO between 1.3 and 1.7, non statistically significant). Prior breast lump is not associated with breast cancer of either pre- or postmenopausal onset, regardless of the definition used.

Overall, in Table 2, when the cutoff for menopause is > 120 months after last menses or > 55 years of age, nulliparity, late age at first live birth and family history of breast cancer appear to be specific risk factors for breast cancer of pre- but not postmenopausal onset. In addition, the RO obtained using the age-based definition of menopause can be either higher or lower than those obtained using the menses-based definition of menopause in a given menopausal group.

Table 3 shows the relationship between nulliparity and breast cancer by 5-year age categories. The highest RO were observed within the age groups below 45

TABLE 2 *Relationship of breast cancer and several reproductive factors according to different definitions of menopause. Maryland and Minnesota, 1973–1975*

Menopausal status Definition	Nulliparity (Yes/No)	Age at first live birth ($\geqslant 25/ < 25$)	Breast cancer in family (Yes/No)	Prior breast lump (Yes/No)
	Adjusted relative odds[a] (95% Confidence limits)			
Premenopausal[b]				
Months after last menses				
$\leqslant 3$	1.5 (0.8–2.8)	2.0 (1.3–3.0)	1.9 (0.9–4.0)	1.2 (0.7–1.9)
$\leqslant 12$	1.6 (0.9–2.9)	2.0 (1.3–3.0)	1.8 (0.9–3.7)	1.2 (0.7–1.9)
$\leqslant 24$	1.7 (0.9–3.1)	2.0 (1.3–3.0)	1.7 (0.9–3.4)	1.2 (0.7–1.8)
$\leqslant 120$	1.7 (1.0–2.8)	1.8 (1.3–2.6)	1.9 (1.1–3.4)	1.4 (0.9–2.0)
Age (years)				
$\leqslant 45$	1.4 (0.8–2.7)	2.0 (1.3–3.0)	2.3 (1.1–4.7)	1.1 (0.7–1.7)
$\leqslant 50$	1.6 (0.9–2.9)	1.8 (1.2–2.6)	2.7 (1.4–5.2)	1.1 (0.7–1.7)
$\leqslant 55$	1.6 (1.0–2.7)	1.9 (1.3–2.7)	2.3 (1.3–4.2)	1.3 (0.9–1.9)
Postmenopausal[b]				
Months after last menses				
> 3	1.0 (0.4–2.1)	1.4 (1.0–2.0)	1.9 (1.0–3.4)	1.3 (0.8–1.9)
> 12	1.0 (0.5–2.1)	1.3 (0.9–1.8)	1.9 (1.0–3.5)	1.2 (0.8–1.9)
> 24	1.0 (0.5–2.1)	1.4 (1.0–2.0)	1.9 (1.0–3.6)	1.2 (0.8–1.9)
> 120	1.1 (0.7–1.8)	1.3 (0.8–2.0)	1.7 (0.8–3.7)	0.8 (0.4–1.5)
Age (years)				
> 45	1.2 (0.5–2.5)	1.4 (1.0–2.0)	1.7 (0.9–3.0)	1.4 (0.9–2.1)
> 50	1.0 (0.5–2.2)	1.5 (1.0–2.2)	1.3 (0.7–2.5)	1.5 (0.9–2.5)
> 55	1.0 (0.5–2.1)	1.3 (0.8–2.0)	1.4 (0.7–2.9)	1.0 (0.6–1.9)

[a] Simultaneously adjusted for age, race, state and education.
[b] All relative odds were statistically homogeneous for pre- and postmenopausal women within each of the seven definitions of menopause.

years but, because of the small number of nulliparous women confidence intervals are very large.

Table 4 shows the difference in accuracy of the three age-based definitions of menopause. Both true positive and false positive per cents tended to be higher in cases than in controls. For each menses-based definition of menopause, the most specific age-based definition of menopause is > 55 years in both cases and controls. But this age-based definition is also the less sensitive. The highest sensitivity for the highest specificity (or lowest false positivity) was observed for the > 50 years old definition. Sensitivity and specificity always differ between cases and controls suggesting that all definitions are associated with differential misclassification bias. The latter do not seem to vary substantially according to definition.

DISCUSSION

The present study shows that the strength of association between breast cancer and some reproductive risk factors may vary according to definition of menopause. Some of the variability is due to statistical imprecision of relative risk estimates because of small sample size or of rare exposure. For example, RO are more often statistically significant in categories with larger numbers of cases and controls such as premenopausal defined '$\leqslant 120$ months after last menses' rather than '$\leqslant 3$ months after last menses'. There is also more variability in RO for a rarer risk factor such as family history of breast cancer. These statistical reasons may explain part of the inconsistencies of published reports, especially for case-control studies smaller than the present one (987 subjects).

TABLE 3 *Relationship between nulliparity and breast cancer within 5-year categories of age. Maryland and Minnesota, 1973–1975*

Age (years)	Nulliparous		Parous		Relative odds (95% confidence limits)
	Cases (No. = 79)	Controls (No. = 79)	Cases (No. = 348)	Controls (No. = 481)	
25–29[b]	4	11	5	32	3.4 (0.4–29.2)
30–34	3	3	18	29	1.5 (0.2–9.2)
35–39	7	4	47	66	1.4 (0.3–5.3)
40–44	10	4	74	96	2.9 (0.8–10.6)
45–49	4	5	42	62	1.1 (0.5–2.5)
50–54	9	7	38	53	1.8 (0.6–5.4)
55–59	8	10	32	46	1.1 (0.4–3.2)
60–64	11	12	37	35	0.9 (0.4–2.2)
65–69	14	11	33	34	1.3 (0.5–3.3)
70–74	9	12	22	28	0.9 (0.3–2.7)

[a] Simultaneously adjusted for race, state and education.
[b] Includes 24 people <25 years.

TABLE 4 *Differences in accuracy of aged-based definition of menopause in cases and controls relative to four menses-based definitions of menopause. Baltimore and Minnesota, 1973–1975*

Definitions of menopause		Cases		Controls	
Age-based	Menses-based (months since last menses)	Sensitivity (% TP)[a]	1-Specificity (%FP)[b]	Sensitivity (%TP)[a]	1-Specificity (%FP)[b]
>45 years	>3	91.3	18.8	79.2	13.5
	>12	90.7	21.0	78.1	14.9
	>24	91.9	24.0	79.9	17.1
	>120	98.6	38.7	89.1	35.0
>50 years	>3	81.7	6.8	66.2	3.6
	>12	82.2	7.7	65.4	4.5
	>24	85.7	8.6	67.7	5.8
	>120	95.9	23.9	80.9	21.3
>55 years	>3	66.3	0.5	51.7	0.9
	>12	67.5	0.5	51.5	0.9
	>24	70.9	0.9	54.1	1.2
	>120	89.8	9.8	74.2	9.8

[a] %TP per cent true positive.
[b] %FP per cent false postive.

Age-based, compared to menses-based definitions of menopause always yielded some degree of misclassification bias. If the bias was of similar magnitude in cases and controls, it would not affect the RO.[31] But, because menopause occurs at a later age in cases than controls, misclassification bias resulting from an age-based definition of menopause must be differential: for any age-based definition, there are more premenopausal women erroneously classified as postmenopausal in cases than in controls. Thus, studies using age-based definitions of menopause are plagued by some degree of differential misclassification bias. In

the present study, ' > 50 years old' was the optimal age-based proxy for menopause, that is, the one with the highest sensitivity for the lowest false positivity.

According to our results, differences in reproductive risk factors between breast cancer of pre- or postmenopausal onset are maximized when menopause is defined as > 120 months after last menses. Using the latter definition of menopause, we found that, nulliparity, late age at first live birth (≥ 25 years) and family history of breast cancer are specific risk factors for breast cancer of premenopausal onset but not for breast cancer of postmenopausal onset. In the only other study that, to our knowledge, used this definition, Wynder et al.[8] also found that late age at first live birth was a specific risk factor for premenopausal women. These findings are biologically plausible since it may take about 10 years for a breast neoplasm to become clinically detectable if initiation occurs after cessation of ovarian hormonal secretion. Another explanation could be that reproductive factors, such as parity, have a short-term protective effect observable in younger but not in older women. However, an effect modification of age is not clear in our data (Table 3) and this interpretation is in sharp contrast with several studies[32-37] showing that nulliparity is protective in younger but deleterious in older women.

In the present study, benign breast disease was not a risk factor for breast cancer. Authors have usually found that fibrocystic but not other forms of benign breast disease were associated with breast cancer,[38] but we did not separate histological types.

In summary, statistical imprecision or differential misclassification bias may explain part of the inconsistent findings of breast cancer studies which used different definitions of menopause. Our results suggest that the effect of reproductive factors on breast cancer aetiology decreases after menopause but this decline takes about 10 years to become observable. We also found that when menstrual history is not available, 50 years of age may be the best proxy definition of menopausal status in populations comparable to that of the present study. Because even a small influence of definition of menopause on weak associations with breast cancer may lead to non-negligible, spurious differences across studies, researchers should check whether their conclusions hold for several definitions of menopause or give a biological rationale for the choice of a particular definition of menopause.

ACKNOWLEDGMENT

REFERENCES

[1] Russo J, Tay K, Russo H. Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Res Treat 1982; 2: 5-73.

[2] Brenner P F, Mishell D R. Menopause. In: Mishell D R, Davajan V (eds). Reproductive endocrinology, infertility and contraception. Philadelphia, F A Davis Co. 1979; pp 161-87.

[3] Hems G. Epidemiological characteristics of breast cancer in the middle and late age. Br J Cancer 1970; 24: 226-34.

[4] De Waard F. The epidemiology of breast cancer; review and prospects. Int J Cancer 1969; 4: 577-86.

[5] Craig T J, Comstock G W, Geiser P B. Epidemiologic comparison of breast cancer patients with early late onset of malignancy and general population controls. JNCI 1974; 53: 1577-81.

[6] Morabia A, Taioli E. Role of reproductive and menstrual history in the etiology of pre- and postmenopausal breast cancer. Aging 1991; 3: 203-09.

[7] Stavraky K, Emmons S. Breast cancer in premenopausal and postmenopausal women. JNCI 1974; 53: 647-54.

[8] Wynder E L, MacCornack F A, Stellman S D. The epidemiology of breast cancer in 785 United States Caucasian women. Cancer 1978; 41: 2341-54.

[9] Cowan L D, Gordis L, Tonascia J A, Jones G S. Breast cancer incidence in women with a history of progesterone deficiency. Am J Epidemiol 1981; 114: 209-17.

[10] Helmrich S P, Shapiro S, Rosenberg L et al. Risk factors for breast cancer. Am J Epidemiol 1983; 117: 35-45.

[11] Valaoras V G, MacMahon G, Trichopoulos D, Polychronopoulou A. Lactation and reproductive histories of breast cancer patients in greater Athens, 1965-67. Int J Cancer 1969; 4: 350-63.

[12] Cusimano R, Dardanoni G, Daranoni L et al. Risk factors of female cancers in Ragusa population (Sicily). 2. Breast cancer. Eur J Epidemiol 1989; 5: 497-506.

[13] Brignone G, Cusimano R, Dardanoni G et al. A case-control study on breast cancer risk factors in a southern European population. Int J Epidemiol 1987; 16: 356-61.

[14] Talamini R, La Vecchia C, Franceschi S et al. Reproductive and hormonal factors and breast cancer in a northern Italian population. Int J Epidemiol 1985; 14: 70-74.

[15] Trichopoulos D, MacMahon B, Cole P. Menopause and breast cancer risk. JNCI 1972; 48: 605-13.

[16] Choi N W, Howe G R, Miller A B et al. An epidemiologic study of breast cancer. Am J Epidemiol 1978; 107: 510-21.

[17] Kelsey J L, Fischer D B, Holford T R et al. Exogenous estrogens and other factors in the epidemiology of breast cancer. JNCI 1981; 67: 327-33.

[18] Brownson R C, Blackwell C W, Pearson D K et al. Risk of breast cancer in relation to cigarette smoking. Arch Intern Med 1988; 148: 140-44.

[19] McDonald J A, Weiss N S, Daling J R, Francis A M, Polissar L. Menopausal estrogen use and the risk of breast cancer. Breast Cancer Res Treat 1986; 7: 193-99.

[20] Ron E, Lunenfield B, Menczer J et al. Cancer incidence in a cohort of infertile women. Am J Epidemiol 1987; 125: 780-90.

[21] Mirra A P, Cole P, MacMahon B. Breast cancer in an area of high parity: Sao Paulo, Brazil. Cancer Res 1971; 31: 77-83.

[22] Staszewski J. Breast cancer and body build. Prev Med 1977; 6: 410-15.

[23] Siskind V, Schofield F, Rice D, Bain C. Breast cancer and breastfeeding: results from an Australian case-control study. Am J Epidemiol 1989; 130: 229-36.

[24] Byers T, Graham S, Rzepka T, Marshall J. Lactation and breast cancer: evidence for a negative association in premenopausal women. Am J Epidemiol 1985; 121: 664-74.

[25] London S J. Colditz G A, Stampfer M J *et al*. Lactation and risk of breast cancer in a cohort of US women. *Am J Epidemiol* 1990; **132**: 17–26.

[26] Paffenbarger R S Jr, Kampert J B, Chang H-G. Characteristics that predict risk of breast cancer before and after the menopause. *Am J Epidemiol* 1980; **112**: 258–68.

[27] McTiernan A, Thomas D B. Evidence for a protective effect of lactation on risk of breast cancer in young women. *Am J Epidemiol* 1986; **124**: 353–58.

[28] Morabia A. *Thyroid hormones and breast cancer*. PhD Dissertation-Baltimore: The Johns Hopkins University, 1989.

[29] Becker R A, Chambers J M, Wilks A R. *The new S language: a programming environment for data analysis and graphics*. Pacific Grove, California: Wadsworth & Brooks, 1988.

[30] Breslow N E, Day N E. *Statistical methods in cancer research: Volume 1: The analysis of case-control studies*. Lyon: IARC, 1980.

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

[32] Janerich D T, Hoff M B. Review and commentary: Evidence for a crossover in breast cancer risk factors. *Am J Epidemiol* 1982; **116**: 737–42.

[33] Lubin J H, Burns P E, Blot W J *et al*. Risk factors for ' . east cancer in women in Northern Alberta, Canada, as related to age at diagnosis. *JNCI* 1982; **68**: 211–17.

[34] Ron E, Lubin F, Wax Y. Re: Evidence for a crossover in breast cancer risk factors. *Am J Epidemiol*, 1984; **119**: 139–40.

[35] Pathak D R, Speizer F E, Willett W C, Rosner B, Lipnick R J. Parity and breast cancer risk: possible effect on age at diagnosis. *Int J Cancer*, 1986; **37**: 21–25.

[36] Negri E, La Vecchia C, Bruzzi P *et al*. Risk factors for breast cancer: pooled results from three Italian case-control studies. *Am J Epidemiol* 1988; **128**: 1207–15.

[37] Thompson W D. Effect modification and the limits of biological inference from epidemiologic data. *J Clin Epidemiol* 1991; **44**: 221–32.

[38] Thomas D B. Epidemiologic and related studies of breast cancer etiology. In: Lilienfeld A M (ed): *Review of cancer epidemiology*. New York: Elsevier; 1980; pp 153–217.

(Revised version received October 1991)