PAPERS

Preterm delivery: effects of socioeconomic factors, psychological stress, smoking, alcohol, and caffeine

Janet L Peacock, J Martin Bland, H Ross Anderson

Abstract

***Objective*—To examine the relation between preterm birth and socioeconomic and psychological factors, smoking, and alcohol and caffeine consumption.**

***Design*—Prospective study of outcome of pregnancy.**

***Setting*—District general hospital in inner London.**

***Participants*—1860 consecutive white women booking for delivery; 1513 women studied after exclusion because of multiple pregnancy and diabetes, refusals, and loss to follow up.**

***Measurements*—Gestational age was determined from ultrasound and maternal dates; preterm birth was defined as less than 37 completed weeks. Independent variables included smoking, alcohol and caffeine consumption, and a range of indicators of socioeconomic status and psychological stress.**

***Main results*—Unifactorial analyses showed that lower social class, less education, single marital status, low income, trouble with "nerves" and depression, help from professional agencies, and little contact with neighbours were all significantly associated with an increased risk of preterm birth. There were no apparent effects of smoking, alcohol, or caffeine on the length of gestation overall, although there was an association between smoking and delivery before 32 weeks. Cluster analysis indicated three subgroups of women delivering preterm: two predominantly of low social status and a third of older women with higher social status who did not smoke. Mean gestational age was highest in the third group.**

***Conclusions*—Adverse social circumstances are associated with preterm birth but smoking is not, apart from an association with very early births. This runs counter to findings for fetal growth (birth weight for gestational age) in this study, where a strong effect of smoking on fetal growth was observed but there was no evidence for any association with psychosocial factors.**

Introduction

Preterm delivery is one of the main causes of perinatal death, neonatal morbidity, and subsequent impairments.[1] Apart from the human costs the economic costs are high—neonatal intensive care has been estimated as costing £600 a day.[2] Early delivery is therefore important both as a clinical problem and a public health issue.

Epidemiological studies can aid prevention by determining risk factors that may be amenable to control on a population basis and by identifying high risk groups that can be targeted by clinical services.[3] Many studies have investigated low birth weight and fetal growth but fewer have looked at preterm delivery.[4] Studies of early delivery have shown varying results. Social class is clearly related to several measures of early child morbidity and mortality,[5] and many studies have reported an association between manual class and preterm birth.[6-9] Other studies that have focused on psychosocial stress have reported increased risk of preterm birth.[6 10-12] Several studies have reported an effect of smoking on length of gestation,[7 13-16] although some have found no relation.[10 17-19] Alcohol has also been implicated,[6 13] but results for caffeine are equivocal.[13 20]

The St George's Hospital birth weight study investigated prospectively relations between outcome of pregnancy and a large number of socioeconomic, psychological, and behavioural variables including smoking and alcohol drinking. The results for birth weight for gestational age (fetal growth), reported previously,[21-23] showed that social factors were not related to fetal growth, but a strong relation was observed between reduced growth and smoking and could not be explained by its correlation with social class. In this paper we examine the relation between the same factors and preterm delivery.

Methods

A consecutive series of 1860 women booking for antenatal care at St George's Hospital between August 1982 and March 1984 were approached for recruitment to a study investigating factors affecting fetal growth. In view of the already well documented effects of ethnic origin on fetal growth and the small size of the various ethnic subgroups, the study was restricted to white women. Other exclusions were poor ability to speak English, age less than 15 years, insulin dependent diabetes mellitus, multiple pregnancy, and presentation later than 24 weeks' gestation. A total of 136 women refused, and 211 failed to complete the study for other reasons (moved, miscarriage, subsequent refusal), leaving a sample of 1513 who had completed interviews one and two and on whom we had outcome data. The numbers with complete data up to 28 and 36 weeks were 1463 and 1433 respectively.

The women were interviewed at four points in pregnancy (booking, and 17, 28, and 36 weeks). Relations were examined between factors measured at the first three interviews and gestational age. Data collected at 36 weeks were available for only a few of those delivering early so were not used in the analysis. Extensive social, behavioural, and psychological data were obtained from the questionnaires, and a detailed obstetric history was taken from the structured hospital record.

Socioeconomic variables included social class (based on the registrar general's classification[24]), education, marital status, cohabitation, housing tenure, and income (see tables). Psychiatric morbidity was assessed with the anxiety and depression scales of the general health questionnaire[25] administered on two occasions. In addition, among questions on health in pregnancy,

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
Janet L Peacock, *lecturer in medical statistics*
J Martin Bland, *reader in medical statistics*
H Ross Anderson, *professor of epidemiology and public health*

Correspondence to:
Dr Peacock.

BMJ 1995;311:531-6

women were asked if "they had suffered any trouble with nerves or depression." Satisfaction and happiness with accommodation, neighbours, neighbourhood, aspects of social support, and feelings about pregnancy were recorded on a four point Likert scale. An inventory of life events modified from Paykel's interview for recent life events was taken at 36 weeks and so is not available for most preterm births.[26]

Smoking was analysed in four categories (non-smokers, former smokers, 1-14 cigarettes a day, ≥15 cigarettes a day) using the reported number of cigarettes smoked at each of two interviews (booking and 28 weeks) and using a combination of the number smoked and brand as indicators of nicotine, carbon monoxide, and tar content.[22] We also calculated the mean of the booking and 28 week reports. Estimated weekly alcohol intake was analysed as grams of alcohol in five groups (0, <20, <50, <100, ≥100). Weekly caffeine intake in milligrams was grouped as none, ≤1400, ≤2800, >2800. Alcohol and caffeine were analysed by using the mean of the first two reports, as for smoking. Data for the booking interview reports only are presented as this allowed all of the women delivering to be included in the analysis and because results were very similar for the combined estimate.

The outcome measure for this analysis was gestational age at delivery. This was calculated from the date of delivery and gestational age at booking based on maternal dates and early ultrasound examination, which was routine at the time of the study. Gestational age has a negatively skewed distribution, which makes it difficult to apply the usual regression methods. We therefore chose to categorise length of gestation. Owing to the small number of very early births, analyses have used the two usual categories of term versus preterm with the cut off of 37 completed weeks for a term birth. However, the relation with smoking was also examined in three groups (<32, 32-36, ≥37 weeks) to test the hypothesis that smoking is related to an excess of very early deliveries.[27] The analyses were done for all births and also for spontaneous and induced births separately. Results are presented for spontaneous births because the induction of birth before 37 weeks is normally associated with clinical problems. Macerated stillbirths and severe congenital malformations were also omitted.

Unifactorial analyses of relations with preterm delivery were performed by using tests for two way tables; χ^2 and Fisher's exact and χ^2 test for trend as appropriate. The results are presented as percentages of preterm births by each category.

The characteristics of the group of women who delivered preterm were investigated with cluster analysis. Other forms of multifactorial analyses that would enable us to isolate particular factors contributing most to the outcome were considered. However, the relevant predictors were highly intercorrelated, making it difficult to produce a unique optimum predictive model. In addition, smoking was related to very early delivery (only) and there were very high preterm birth rates in some of the most deprived groups. Cluster analysis allowed us to investigate this by producing a profile of women at risk.

Since most of the variables of interest are categorical, the method adopted was association analysis.[28] Each variable was regrouped into two categories chosen to give as near equal numbers in the groups as possible, consistent with the nature of the groups. For example, marital status was regrouped as "married" and "not married." The analysis first finds the single variable most closely related to all the others by cross tabulating each pair of variables. The variable with the greatest sum of χ^2 statistics is the one that is most closely related to all the others. The sample is then split in two using the two categories of this variable. The process is repeated for these two groups and continues until the sums of χ^2 statistics are all below the conventional 5% significance level.[29] This process groups individual subjects who are similar in terms of the variables used.

Statistical analyses were done using SAS[30] and our own software.

Results

UNIFACTORIAL ANALYSES

In all, 7·5% of births (113/1513) were preterm. Of these, 14% (13/96) were induced and 86% (83/96; onset of labour was missing for 17/113) were spontaneous labours. The expected relation between preterm birth and birth weight was seen, and there was a tendency for these early births to be associated with retarded growth, although this was not significant. Sex of the baby was not significantly associated with gestational age. Teenage mothers were more likely to deliver preterm but there was no significant linear trend with age (table I). Maternal height was not related to gestational age. Primiparous women had more early deliveries, but this was not significant.

There was twofold variation in rates of preterm birth by five main indicators of socioeconomic disadvantage: social class, education, marital status, housing tenure, and income (table II). Most of these associations were significant. We only tested for overall association or trend, if appropriate, and did not test differences between observed extreme groups. Altogether 26 "social" factors were tested, of which five were statistically significant at the 5% level (table II), although the twofold differences were often seen between extreme groups (data not presented).

There was no evidence of a relation between anxiety score and preterm birth (table III), nor was there any trend in early delivery with levels of depression. Women who reported "trouble with nerves and depression" had a nearly twofold excess of preterm births but this was not significant. Indicators of social support showed similar twofold variations with preterm delivery, some of which were significant (table III). Some studies have found associations with adverse life events. We could not use our data on life events because they were collected at 36 weeks' gestation. However, an increase in life events was associated with single marital status, rented accommodation, low

TABLE I—*Characteristics of the study population (n=1513)*

Factor	% (No) preterm*	P value
Maternal age:		
15-19	10·1 (10/99)	0·33 (trend)
20-24	8·2 (31/378)	
25-29	6·6 (35/532)	
30-34	7·9 (28/356)	
≥35	6·1 (9/148)	
Maternal height (cm):		
142-155	5·8 (11/190)	0·25 (trend)
156-160	7·8 (29/373)	
161-165	6·5 (32/489)	
166-170	9·0 (27/299)	
≥171	8·6 (14/162)	
Parity:		
Primiparous	8·2 (63/765)	0·29
Multiparous	6·7 (50/748)	
Onset of labour:		
Spontaneous	6·9 (83/1201)	0·49
Induced	5·4 (13/239)	
Sex of baby:		
Boy	7·3 (55/755)	0·86
Girl	7·6 (58/758)	
Birth weight (g):		
<2500	66·7 (52/78)	<0·0001
≥2500	4·2 (61/1435)	
Growth retardation:		
<10th centile	10·7 (16/150)	0·16
≥10th centile	7·1 (97/1363)	

*Totals vary because of missing data.

TABLE II—*Social variables and spontaneous preterm birth (n=1201)*

Factor	% (No) preterm*	P value
Social class†:		
I	4·4 (4/90)	0·02 (trend)
II	4·1 (13/318)	
IIIN	6·4 (7/109)	
IIIM	9·3 (43/463)	
IV	6·4 (7/109)	
V	8·2 (4/49)	
Education:		
Minimum	9·6 (55/575)	0·001
More than minimum	4·5 (28/622)	
Marital status:		
Married	6·2 (59/960)	0·03
Divorced or separated or widowed	4·4 (2/45)	
Single	11·2 (22/196)	
Cohabitation:		
Yes	6·6 (72/1100)	0·15
No	10·9 (11/101)	
Housing tenure:		
Owner	5·4 (38/702)	0·07
Council rent	10·5 (22/210)	
Private rent	5·5 (8/145)	
Living with parents	3·5 (2/58)	
Other	9·7 (3/31)	
Monthly income‡:		
0-209	12·5 (8/64)	0·04 (trend)
210-319	7·1 (6/84)	
320-429	7·7 (7/91)	
430-539	7·6 (12/158)	
540-644	6·5 (9/139)	
645-754	8·0 (11/138)	
755-859	3·6 (5/133)	
≥860	5·3 (16/300)	
Managing on income:		
Not difficult	6·8 (47/687)	0·76 (trend)
Slightly difficult	6·7 (19/284)	
Fairly difficult	8·3 (11/132)	
Very difficult	6·9 (6/87)	
Phone:		
Yes	6·0 (59/990)	0·22
No	8·9 (14/157)	
Satisfaction with accommodation:		
Very satisfied	5·8 (37/638)	0·67 (trend)
Fairly satisfied	7·3 (27/370)	
Fairly unsatisfied	7·9 (6/76)	
Very unsatisfied	4·8 (3/62)	
Planned pregnancy:		
Yes	6·2 (53/849)	0·89
No	6·7 (20/299)	
No of adults in household:		
1	8·1 (3/37)	0·06 (trend)
2	6·2 (63/1012)	
≥3	11·6 (17/147)	

Other variables assessed and not significant: employment; mother's social class; affording clothes, food, heating, rent; housing, number of rooms; household amenities; satisfaction with neighbourhood; mother's, partner's feelings about pregnancy; termination considered.
Other variables assessed and significant: Qualifications of mother.
*Totals vary because of missing data.
†Mother's father's social class.
‡Net monthly income (£) in 1982-4.

income, and less education in women who gave birth at term.

Smoking did not have a negative influence on gestational age overall (table IV). Non-smokers had a higher rate than former smokers and current smokers. In addition, we examined the hypothesis that smoking is related to very early delivery (<32 weeks) and found an association with very early birth (relative risk at <32 weeks=1·95 (95% confidence interval 1·30 to 2·93); at 32-36 weeks=0·82 (0·56 to 1·21); at ≥37 weeks=1·00).

Alcohol consumption and gestational age were not significantly related. Non-drinkers had the highest risk of preterm birth, and the heaviest drinkers had the lowest. Caffeine intake was not associated with early birth. Those with a low caffeine intake had the highest rates. These lack of associations held for consumption at each interval in pregnancy and for mean intake.

CLUSTER ANALYSIS

The unifactorial analyses above suggested a relation between adverse social circumstances and preterm birth. Smoking was related to very early delivery. In addition, there were very high rates in some of the most deprived groups. Cluster analysis was used to investigate this among all the women spontaneously delivering preterm. The variables included were age (15-24 *v* ≥25), social class (non-manual *v* manual), education, marital status (married *v* not married), income (below *v* above median), "nerves," and smoking. The analysis resulted in three clusters described below.

Age was the variable most closely associated with all the others and so the group was split into younger and older women (table V). The lower age group did not divide further; the upper age group was split by smoking status.

Cluster 1: younger women—These 36 women were of predominantly manual social class, with lower income and minimum education; half were married; one third reported "trouble with nerves and depression." The mean gestational age in this group was 34·4 weeks.

Cluster 2: older women, smokers—These 11 women were all of manual social class, mostly with lower income and minimum education and married; one fifth reported "trouble with nerves and depression." The mean gestational age was 33·9 weeks.

TABLE III—*Psychological factors, social support, and spontaneous preterm birth (n=1201)*

Factor	% (No) preterm*	P value
Psychological factors		
Anxiety score:†		
0-1	7·3 (27/368)	0·76 (trend)
2-3	6·6 (18/273)	
4-6	6·5 (18/276)	
≥7	6·8 (19/280)	
Depression score:†		
0	6·2 (53/859)	0·34 (trend)
1	9·7 (14/144)	
2-3	7·5 (8/107)	
≥4	8·1 (7/86)	
"Trouble with nerves and depression":‡		
No	6·3 (66/1042)	0·06
Yes	10·8 (17/158)	
Social support		
Help from professional agencies:		
None	5·9 (62/1053)	0·05
Help received	11·6 (11/95)	
Confidant:		
Husband or partner	2·2 (1/45)	0·56
Husband and other	5·8 (23/400)	
Other	7·1 (45/636)	
None	6·4 (4/63)	
Contact with neighbours:		
Daily	5·2 (23/445)	0·01 (trend)
Weekly	5·6 (26/467)	
Monthly	6·7 (4/60)	
Less than monthly	16·2 (6/37)	
Never	10·4 (14/134)	

Other variables assessed and not significant: contact with relatives, friends; "trouble with nerves and depression" at 28 weeks.
Other variables assessed and significant: "Get on with neighbours."
*Totals vary because of missing data.
†Assessed at booking; maximum score 21.
‡Assessed at booking.

TABLE IV—*Smoking and alcohol and caffeine consumption at booking, and spontaneous preterm birth (n=1201)*

Factor	% (No) preterm*	P value
Smoking habit:		
Non-smoker	8·4 (26/309)	0·45 (trend)
Former smoker	6·0 (29/482)	
1-14 cigarettes a day	7·3 (21/289)	
≥15 cigarettes a day	5·8 (7/120)	
Alcohol consumption (g/week):		
None	8·4 (50/597)	0·21 (trend)
1-19	4·6 (14/304)	
20-49	6·2 (12/194)	
50-99	8·7 (6/69)	
≥100	2·9 (1/34)	
Caffeine intake (mg/week):		
None	7·1 (2/28)	0·56 (trend)
1-1400	8·2 (25/305)	
1401-2800	6·2 (29/470)	
≥2801	6·9 (27/393)	

Other variables assessed and not significant: smoking, alcohol, and caffeine at 28 and 36 weeks gestation; mean smoking, alcohol, and caffeine; total smoking assessed using both brand and number smoked[22]; passive smoking.
*Totals vary because of missing data.

TABLE V—χ^2 *values for cluster analysis*

	Age	Social class	Education	Marital status	Income	Nerves	Smoking	Total
				All 7 variables				
Age		3·29	4·73	8·85	10·21	7·92	4·16	39·16*
Social class	3·29		7·17	2·68	1·13	0·00	4·26	18·53
Education	4·73	7·17		3·39	3·35	0·51	5·90	25·05
Marital status	8·85	2·68	3·39		8·35	4·62	7·66	35·55
Income	10·21	1·13	3·36	8·35		3·80	5·44	32·29
"Nerves"	7·92	0·00	0·51	4·62	3·80		2·53	19·38
Smoking	4·16	4·26	5·90	7·66	5·44	2·53		29·95
*χ^2=39·16, df=6, p<0·001—split by age								
				Split by age—younger women				
Social class			1·84	0·00	0·00	0·23	0·00	2·07
Education		1·84		0·46	1·12	0·00	0·46	3·88
Marital status		0·00	0·46		1·60	0·06	2·73	4·85†
Income		0·00	1·12	1·60		0·00	0·00	2·72
"Nerves"		0·23	0·00	0·06	0·00		0·06	0·35
Smoking		0·00	0·46	2·73	0·00	0·06		3·25
†χ^2=4·85, df=5, p>0·25—split no more								
				Split by age—older women				
Social class			2·34	1·19	0·58	0·00	6·56	2·07
Education		2·34		0·27	0·11	0·00	2·80	5·52
Marital status		1·19	0·27		1·03	1·76	0·70	4·95
Income		0·58	0·11	1·03		2·67	5·37	9·76
"Nerves"		0·00	0·00	1·76	2·67		0·49	4·92
Smoking		6·56	2·80	0·70	5·37	0·49		15·92‡
‡χ^2=15·92, df=5, p<0·01—split by smoking								
				Split further by smoking—older smokers				
Social class			0·00	0·00	0·00	0·00		0·00
Education		0·00		0·01	0·01	0·00		0·02
Marital status		0·00	0·01		0·23	2·81		3·05§
Income		0·00	0·01	0·23		0·01		0·25
"Nerves"		0·00	0·01	2·81	0·01			2·83
§χ^2=3·05, df=4, p>0·50—split no more								
				Split further by smoking—older non-smokers				
Social class			0·71	0·35	0·00	0·47		1·53‖
Education		0·71		0·00	0·00	0·42		1·13
Marital status		0·35	0·00		0·00	0·00		0·35
Income		0·00	0·00	0·00		0·29		0·29
"Nerves"		0·47	0·42	0·00	0·29			1·18
‖χ^2=1·53, df=4, p>0·75—split no more								

TABLE VI—*Factors associated with gestational age and fetal growth in the St George's birth weight study*

Factor	Associations with gestational age	Associations with fetal growth
Smoking	Higher risk among <32 weeks; no relation 32-36 weeks	Strong relation—not explained by social class
Alcohol consumption	No relation	Relation among smokers only
Caffeine intake	No relation	Relation among smokers only
Social class	Higher risk among low class not explained by smoking	Relation explained by smoking
Education	Higher risk among less educated	Relation explained by smoking
Marital status	Higher risk among single women	No relation
Income	Higher risk among low income	No relation
Psychological factors	No relation	No relation
Social support	Higher risk among those in contact with professional agencies and those with little contact with neighbours	No relation

Cluster 3: older women, non-smokers—These 36 women had predominantly higher income, were married, half with minimum education and half manual social class. Few reported "trouble with nerves and depression." The mean gestational age was 35·0 weeks.

Discussion

This study has provided some evidence for an association between preterm birth and several socioeconomic factors (social class, education, marital status, income, help from professional agencies, contact with neighbours). There was no evidence for a relation of gestational age overall with smoking or alcohol or caffeine intake. However this study confirms Meyer's hypothesis that smoking is related to very early delivery,[27] with a twofold increase in risk of delivery before 32 weeks. This hypothesis was further supported by the results of the cluster analysis of women delivering early, which showed that mean gestational age was lowest among the cluster of smokers over age 24. Smoking does not seem to be a factor among preterm births after 32 weeks' gestation. The cluster analysis indicated the presence of three subgroups of women delivering preterm: two clusters were predominantly of low social status, and the third cluster comprised older women with higher social status who did not smoke. Mean gestational age was highest in this third group, although differences were small.

The St George's data has been carefully analysed to look at the two outcomes, fetal growth and gestational age, separately (table VI). With respect to fetal growth, there was little evidence of a relation with social class, which is in contrast with the relation observed for gestational age. Conversely, smoking was strongly related to reduced fetal growth, but was associated with gestational age only among the women who delivered before 32 weeks (12/83; 14%). When fetal growth was considered, the effect of smoking could not be explained by social class, but the effect of social class was explained by smoking. With respect to gestational age, the association with social class could not be explained by smoking. Hence the apparent effect of social factors on gestational age does not seem to be mediated through smoking. These findings provide further evidence for different aetiologies for fetal growth retardation and preterm birth.[31]

The observed effect of socioeconomic factors may operate in several ways to reduce gestational age. It could be explained by qualitative and quantitative differences in nutrition[8] or by the observed correlation of socioeconomic factors with adverse life events.[12] Although this study was unable to measure life events directly, a correlation was observed between life events and housing tenure, marital status, income, and education. The direct mechanism through which stress might cause early labour is through catecholamine release and changes in other hormone concentrations.[32] There is also increasing evidence that lower genital tract infection is the cause for some preterm labours.[33] Previous analyses of our data have shown a higher prevalence of symptoms and health problems in pregnancy among manual social classes,[34] although our sample was too small to examine the association with preterm birth. It is therefore possible that some women deliver preterm because of an increased prevalence of infection. More clinical research is needed to further understand the causal pathways.

The main analysis in this study included only spontaneous births. We hypothesised that the observed relations between early delivery and social status might be due to a "selection out" of excess induced births among more privileged women. This hypothesis was dismissed when we found that induction rates at term did not vary by social class or education. Women who reported trouble with nerves and depression were less likely to have induced births than those reporting no problems for both preterm and term births. Therefore the observed relation is not explained by a selection effect.

Because many tests of significance were performed in these analyses, the significant results could be type I errors. In addition, many variables were intercorrelated. Thirty six tests were performed on socioeconomic variables, of which eight were significant at the 5% level. This study showed some large differences between subgroups which were not significant (this was investigated further using the cluster analysis). This "lack" of significant results could be due to low power. Given the sample size of 1200, power 0·9, significance level 0·05, and two equal size groups, a difference in the preterm birth rate of about 5% could be detected. If the subgroups were not of equal size (as they were not here), a difference of 7% was detectable.

Birth weight is routinely recorded and is regarded as an important indicator of reproductive health either as

Key messages

- This study shows that five main indicators of socioeconomic disadvantage are associated with twofold variation in rates of preterm birth
- Smoking is associated only with increased risk of very early delivery; there is no excess risk after 32 weeks' gestation
- This contrasts with the findings for fetal growth, for which there was no association with social factors but a strong relation with smoking
- Preterm birth and fetal growth have different aetiologies
- Associations with the outcome birth weight should be interpreted in the light of its two components: fetal growth and gestational age

a distribution or as a proportion of low birth weight. The two components of birth weight are fetal growth and gestational age. We have shown that the risk factors for these two components are different. This has implications for the interpretation of variations in birth weight between populations and for public health policy. Further, an understanding of factors affecting these two components of birth weight may provide clues about possible mechanisms for observed associations between birth weight and later social disadvantage.[35]

We thank Professors R R Trussell and G V P Chamberlain and the clinic staff for facilitating, and Dr O G Brooke for initiating the study, Malcolm Stewart for his help in organising the data collection, all the interviewers, and the pregnant women for participating.

Funding: Data collection and previous analyses only (already published[21]) were funded by a consortium of American Tobacco Companies.

Conflict of interest: None.

1 Bakketeig LS, Hoffman HJ. Epidemiology of preterm birth. In: Elder MG, Hendricks CH, eds. *Preterm labor.* London: Butterworths, 1981:17-46.
2 Griffin J. *Born too soon.* London: Office of Health Economics, 1993.
3 Rose G. *The strategy of preventive medicine.* Oxford: Oxford Medical Publications, 1992.
4 Berendes H. Implications for future avenues of epidemiological research. In: *Proceedings of the international collaborative effort on perinatal and infant mortality.* Vol III. Rockeville, MD: US Dept of Health and Human Services, 1992: VIII-11.
5 Rutter DR, Quine L. Inequalities in pregnancy outcome: a review of psychosocial and behavioural mediators. *Soc Sci Med* 1990;30:553-68.
6 Berkowitz GS. An epidemiologic study of preterm delivery. *Am J Epidemiol* 1981;113:81-92.
7 Fedrick J, Anderson ABM. Factors associated with spontaneous pre-term birth. *Br J Obstet Gynaecol* 1976;83:342-50.
8 Garn SM, Shaw HA, McCabe KD. Effects of socioeconomic status and race on weight-defined and gestational prematurity in the United States. In: Reed DM, Stanley FJ, eds. *The epidemiology of prematurity.* Baltimore: Urban and Schwarzenberg, 1977:127-43.
9 Pickering RM, Deeks JJ. Risks of delivery during the 20th to 36th week of gestation. *Int J Epidemiol* 1991;20:456-66.
10 Abernathy JR, Greenberg BG, Bradley Wells H, Frazier TM. Smoking as an independent variable in a multiple regression analysis upon birth weight and gestation. *Am J Public Health* 1966;56:626-33.
11 Hedegaard M, Henriksen TB, Sabroe S, Secher NJ. Psychological distress in pregnancy and preterm delivery. *BMJ* 1993;307:234-9.
12 Newton RW, Webster PAC, Binu PS, Maskrey N, Phillips AB. Psychosocial stress in pregnancy and its relation to the onset of premature labour. *BMJ* 1979;ii:411-3.
13 McDonald AD, Armstrong BG, Sloan M. Cigarette, alcohol and coffee consumption and prematurity. *Am J Public Health* 1992;82:87-90.
14 Hartikainen-Sorri A, Sorri M. Occupational and sociomedical factors in preterm birth. *Obstet Gynecol* 1989;74:13-6.
15 Meyer MB, Jonas BS, Tonascia JA. Perinatal events associated with maternal smoking during pregnancy. *Am J Epidemiol* 1976;103:464-76.
16 McIntosh ID. Smoking and pregnancy. II. Offspring risks. *Public Health Rev* 1984;12:29-63.
17 Berkowitz GS, Kasl SV. The role of psychosocial factors in spontaneous preterm delivery. *J Psychosomatic Res* 1983;27:282-90.
18 Roberts WE, Morrison JC, Hamer C, Wiser WL. The incidence of preterm labor and specific risk factors. *Obstet Gynecol* 1990;76:85S.
19 Stein A, Campbell EA, Day A, McPherson K. Social adversity, low birth weight, and preterm delivery. *BMJ* 1987;295:291-3.
20 Fortier I, Maroux S, Beaulac-Baillargeon L. Relation of caffeine intake during pregnancy to intrauterine growth retardation and preterm birth. *Am J Epidemiol* 1993;137:931-40.
21 Brooke OG, Anderson HR, Bland JM, Peacock JL, Stewart CM. Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. *BMJ* 1989;298:795-801.
22 Peacock JL, Bland JM, Anderson HR, Brooke OG. Cigarette smoking and birthweight: type of cigarette smoked and a possible threshold effect. *Int J Epidemiol* 1991;20:405-12.
23 Peacock JL, Bland JM, Anderson HR. Effects on birthweight of alcohol and caffeine consumption in smoking women. *J Epidemiol Community Health* 1991;45:159-63.
24 Office of Population Censuses and Surveys. *Classification of occupations.* London: HMSO, 1980.
25 Goldberg DP, Hillier VF. A scale version of the general health questionnaire. *Psychol Med* 1979;9:139-45.
26 Paykel ES, Myers KJ, Dienelt MN, Lerman GL, Lindenthal JJ, Pepper MP. Life events and depression: a controlled study. *Arch Gen Psychiatry* 1969;21:753-60.
27 Meyer MB. Effects of maternal smoking and altitude on birthweight and gestation. In: Reed DM, Stanley FJ, eds. *The epidemiology of prematurity.* Baltimore: Urban and Swartzenberg, 1977:81-104.
28 Everitt B. *Cluster analysis.* 2nd ed. London: Heinemann Educational, 1980.
29 MacNaughton-Smith P. Some statistical and other numerical techniques for classifying individuals. London: HMSO, 1965.
30 SAS Institute. *SAS users guide.* Cary, NC: SAS Institute, 1985.
31 Arbuckle RE, Sherman GJ. Comparison of the risk factors for preterm delivery and interuterine growth retardation. *Paediatric and Perinatal Epidemiology* 1989;3:115-29.
32 Main DM. The epidemiology of preterm birth. *Clin Obstet Gynecol* 1988;31: 521-32.
33 Hay PE, Lamont RF, Tayor-Robinson D, Morgan DJ, Ison C. Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage. *BMJ* 1994;308:295-8.
34 Meyer LM, Peacock JL, Bland JM, Anderson HR. Symptoms and health problems in pregnancy: their association with social factors, smoking, alcohol, caffeine and attitude to pregnancy. *Paediatric and Perinatal Epidemiology* 1994;8:145-55.
35 Bartley M, Power C, Blane D, Davey Smith G, Shipley M. Birth weight and later socioeconomic disadvantage: evidence from the 1958 British cohort study. *BMJ* 1994;309:1475-9.

(Accepted 15 June 1995)

Commentary: Classification and cluster analysis

B S Everitt

Institute of Psychiatry, London SE5 8AL
B S Everitt, *professor of statistics in behavioural science*

One of the most basic abilities of living creatures involves the grouping of similar objects to produce a classification. As well as being a basic human conceptual activity, classification is also fundamental to most branches of science. In chemistry, for example, the classification of the elements in the periodic table has had a profound impact on the understanding of the structure of the atom. Classification in medicine is equally important, with the classification of diseases being of primary concern as the basis for investigations of aetiology and treatment.

Statistical techniques for classification are essentially of two types. Members of the first type are used to construct a (hopefully) sensible and informative classification of an initially unclassified set of data; these are known as cluster analysis methods. The information on which the derived classification is based is generally a set of variable values recorded for each patient or individual in the investigation, and clusters are constructed so that individuals within clusters are similar with respect to their variable values and different from individuals in other clusters. Paykel and Rassaby, for example, studied 236 people who had attempted suicide presenting at the main emergency service of one American city.[1] Each patient was described by 14 variables including age, number of previous suicide attempts, and severity of depression. A number of clustering methods were applied to the data and a final classification with three groups was produced which appeared potentially valuable as a