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 published²¹) 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. Rockville, 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 Hedegeard 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;iii: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-Bailargeon 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, Taylor-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

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

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

Data can sometimes be clustered simply by visual examination

basis for future studies into the causes and possible treatment of attempted suicide. (The second set of statistical techniques concerned with classification is known as discriminant or assignment methods. Here the classification scheme is known a priori and the problem is how to devise rules for allocating unclassified individuals to one or other of the known classes.)

Many methods have been suggested for constructing a set of clusters from numerical data, ranging from the visual examination of simple scatterplots of data (fig) to the fitting of complex statistical models. A comprehensive review is given in Everitt.² Among these methods, no single method is best in all situations and for all types of data. Different methods of cluster analysis can (and probably will) produce different solutions when applied to the same set of data. As a consequence, clustering techniques require great care in their practical application if misleading and artificial solutions are to be avoided.

The set of clusters produced by a clustering method can serve a variety of purposes. At one level they may simply represent a convenient method for organising a large dataset so that information may be retrieved more efficiently. In general, however, investigators are likely to be looking for their solutions to represent some more fundamental feature of the data, which is why the caveats offered above are necessary to prevent artefact becoming theory.

Peacock, Bland, and Anderson use cluster analysis in an attempt to identify women at risk of preterm birth. Seven binary variables are used to characterise each woman in the study who spontaneously delivered preterm. The method of cluster analysis used, association analysis, is monothetic—that is, division of the data into clusters is based on the possession or otherwise of a single specified characteristic. Division variables are chosen in a particular way, as described in the paper. The authors report a solution with three clusters, although the first two of these seem to be similar.

The use of cluster analysis in such a study is certainly a novel way of attacking the problem of identifying subsets of risk factors among highly correlated explanatory variables. It is unfortunate that the authors used a rather outdated clustering method, one based on divisions involving a single variable at each stage. It would be interesting to compare the results obtained with those given by a method such as latent class analysis,² which is also specifically designed for binary variables, but is underlined by a relatively sensible statistical model. Nevertheless, Peacock, Bland, and Anderson are to be congratulated for alerting other investigators to the advantages of using cluster analysis in this type of study.

- 1 Paykel ES, Rassaby E. Classification of suicide attempters by cluster analysis. *Br J Psychiatry* 1978;133:45-52.
- 2 Everitt BS. *Cluster analysis*. 3rd ed. London: Edward Arnold, 1993.

Optimal polymerase chain reaction amplification for preimplantation diagnosis in cystic fibrosis ($\Delta F508$)

Ke-Hui Cui, Eric A Haan, Ling-Jia Wang, Colin D Matthews

Abstract

Objective—To evaluate direct polymerase chain reaction amplification of mutation on single embryo cells for the routine preimplantation diagnosis of cystic fibrosis.

Design—Direct polymerase chain reaction amplification of mutation was performed to identify the cystic fibrosis $\Delta F508$ mutation in human blood DNA, single lymphocytes, embryos, and embryo cells obtained by biopsy. Preimplantation diagnosis was performed for a couple who were heterozygous carriers of the $\Delta F508$ mutation.

Setting—Laboratory for preimplantation diagnosis in a reproductive medicine unit.

Main outcome measure—Correct diagnosis of homozygous normal, heterozygous, and homozygous abnormal DNA of the cystic fibrosis $\Delta F508$ mutation.

Results—45 blood samples (18 homozygous normal, 17 heterozygous, and 10 homozygous abnormal) and 204 single lymphocytes from known sources showed 100% amplification and were diagnosed correctly. 17 human embryos and 52 normal nucleated embryo cells obtained by single cell embryo biopsy also showed 100% amplification. After a miscarriage of the initial pregnancy (diagnosed at preimplantation to be homozygous normal) in the heterozygous carrier couple, fetal tissue was confirmed to be homozygous normal.

Conclusion—Direct polymerase chain reaction amplification of mutation is a simple, fast, reliable test for the common cystic fibrosis mutation ($\Delta F508$) in blood DNA and single cells and should be applicable to routine programmes of general

screening, maternal blood examination, and preimplantation diagnosis.

Introduction

Cystic fibrosis is the most prevalent severe inherited disease in white people with an incidence of about one in 2500 births and a carrier frequency of about one in 25.¹ The disorder is caused by over 200 mutations in a gene called the cystic fibrosis transmembrane regulator, which is on chromosome 7q31.² The commonest mutation ($\Delta F508$) is the deletion of an amino acid (phenylalanine) at position 508 of the gene,³ which is responsible for 70% of cystic fibrosis cases.

Preimplantation diagnosis is a technique of prenatal diagnosis that avoids medical abortion. Precision of diagnosis—that is, as close to 100% accuracy as possible—is crucial⁴ and is particularly demanding technically, given that a single copy of the gene must be amplified successfully from a single embryo cell obtained by single cell biopsy. Precision of preimplantation diagnosis has been achieved in sex determination by polymerase chain reaction.^{5,6} Polymerase chain reaction is a method by which a targeted gene is greatly amplified by a polymerase enzyme with two short leading segments of the gene (primers) and other reagents by many cycles of suitable temperature changes in a thermal cycler. Precision of diagnosis is wholly dependent on the method used. Any inappropriate method is easily detected by basic experiments⁷⁻¹¹ and will be shown directly by misdiagnosis.¹² A technique that achieves near 100% precision of diagnosis would provide a scientific basis for routine clinical practice in preimplantation diagnosis.¹³⁻¹⁶

Department of Obstetrics and Gynaecology, University of Adelaide, Queen Elizabeth Hospital, Woodville, Adelaide, SA 5011, Australia
Ke-Hui Cui, scientist in charge of preimplantation diagnosis
Ling-Jia Wang, scientist
Colin D Matthews, professor

Department of Medical Genetics, Women's and Children's Hospital, North Adelaide, SA 5006
Eric A Haan, director of department

Correspondence to: Dr Cui.

BMJ 1995;311:536-40