# Representativeness of samples from general practice lists in epidemiological studies: case-control study

Alexandra G Smith, Nicola T Fear, Graham R Law, Eve Roman

Epidemiology and Genetics Unit, University of Leeds, Leeds LS2 9LN
A G Smith *research fellow*
N T Fear *senior research fellow*
G R Law *senior research fellow*
E Roman *professor of epidemiology*

Correspondence to: A G Smith alex.smith@egu.leeds.ac.uk

*BMJ* 2004;328:932

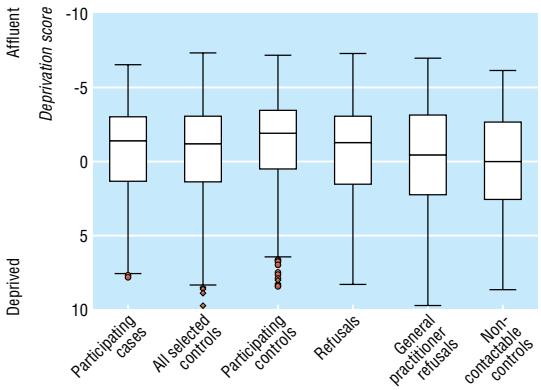
Ethical constraints often prevent epidemiological studies from evaluating the impact of non-participation. Particular problems may arise when subjects fail to respond to an approach by researchers or when they cannot be contacted because of inaccurate contact details or a doctor's refusal to give permission for their patient to be approached. If these subjects differ from those subjects who agree or decline to participate then the validity and generalisability of the study may be compromised. We investigated these issues in a case-control study of acute leukaemia in England.

## Participants, methods, and results

The details of the study have been previously described.[1] For each case, 10 people were randomly chosen from the case's general practice list, matched on sex and year of birth. With the general practitioner's consent, the first two controls identified were sent a letter explaining the study and inviting them to participate. If no reply was received within two weeks, the subject was telephoned, and if no reply (or a negative reply) had been received within a month from the initial contact date then the next control on the list was approached. This continued until two controls per case had agreed to participate. All subjects, regardless of participation, were assigned a Townsend material deprivation score based on area of residence[2] at the enumeration district level, which contained aggregated census information from about 200 households.

Overall, 838 cases participated, and 3540 controls were selected, of whom 1658 participated (47%), 854 (24%) declined, 715 (20%) could not be contacted at the address held, and 313 (9%) could not be contacted because their general practitioner refused to give permission. The main reason that patients gave for not participating was because they did not have the time to be interviewed. General practitioners refused permission for their patient to be approached largely because of the patient's family or personal circumstances such as illness or social problems. Unfortunately, no further information about those who could not be contacted was available.

The figure shows the mean deprivation score for the areas in which cases and controls lived, according to participation status. Although the selected controls lived in areas of similar material wealth to their corresponding cases, the controls who participated differed markedly from those who did not. Furthermore, we found significant differences ($P<0.05$) between the non-participating groups. Those who could not be contacted tended to live in the most deprived areas, followed by those whose GP refused contact and those who were contacted but declined to participate. The deprivation distributions between the subgroups may seem similar, especially compared with the possible range of deprivation scores for England and Wales ($-8$ to 12). However, as the controls were selected from the same general practice surgery as their corresponding case, the subjects were effectively matched on area of residence as patients in the United Kingdom usually live in a defined catchment area around the practice.




Distribution of deprivation score by participation status and reason for non-participation

## Comment

Given the well established link between socioeconomic status and health,[3] findings based on subjects who participate in epidemiological studies that have used GP registers to identify subjects may not be generalisable to the population as a whole.

Sampling from GP registers, however, is a popular and convenient way of finding subjects. Indeed, the sampling frame of UK Biobank (www.ukbiobank.ac.uk) is predicated on volunteers selected from GP lists. Although for some investigations this may not be critical, for others the under-representation of certain minority groups may well result in false negative findings and biased estimates of risk.

The epidemiology and genetics unit is moving to the University of York (YO10 5SD) on 1 April 2004. We thank E V Willett and G J Dovey for their comments on the manuscript.

AGS and ER initiated the research. AGS also did the data analysis and is the guarantor. All authors helped to interpret the results and write the paper.

Funding: Leukaemia Research Fund.

Competing interests: None declared.

Ethical approval: Local ethical research committees.

1 Kane EV, Roman E, Cartwright R, Parker J, Morgan G. Tobacco and the risk of acute leukaemia in adults. *Br J Cancer* 1999;81:1228-33.
2 Townsend P, Phillimore P, Beattie A. *Health and deprivation: inequalities in the north*. London: Croom Helm, 1988.
3 Department of Health and Social Security. *Inequalities in health*. London: DHSS, 1980. (Report of working group chaired by Sir Douglas Black.)

*(Accepted 26 November 2003)*

doi 10.1136/bmj.38029.672662.AE

*This article was posted on bmj.com on 27 February 2004: http://bmj.com/cgi/doi/10.1136/bmj.38029.672662.AE*