Law, GR; Smith, AG; Roman, E; United Kingdom Childhood Cancer Study Investigators; (2002) The importance of full participation: lessons from a national case-control study. British journal of cancer, 86 (3). pp. 350-355. ISSN 0007-0920 DOI: https://doi.org/10.1038/sj.bjc.6600092

Downloaded from: http://researchonline.lshtm.ac.uk/id/eprint/15764/

DOI: https://doi.org/10.1038/sj.bjc.6600092

Usage Guidelines:

Please refer to usage guidelines at https://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
The importance of full participation: lessons from a national case–control study

Differential participation between cases and controls can lead to biased estimates of risk. However, the effects of participation are often ignored. We report a detailed analysis of locations of residence for participants and non-participants in a large, national case–control study of childhood cancer in Great Britain, using the 1991 census. The initial selection of 7669 controls, taken from lists of those registered with a General Practitioner, was representative of the British population in respect to an areal-based index of material deprivation. However, parents of controls agreeing to participate were living in more affluent areas than initially selected controls and their matched 3838 cases. The three components of the deprivation index, persons unemployed, households not owning a car or their home were similarly associated with participation. Other census characteristics, such as proportion of flat dwellers and centrally heated households were also associated with control participation. Population density of the local area was not different between participating controls and their matched cases. However, initially selected controls lived in more urban areas than their cases. Such differences are not unique to this study, as they are an inevitable consequence of incomplete participation. The implications of these differences are discussed, in relation to the difficulty this imposes in the interpretation of studies of disease aetiology.

MATERIALS AND METHODS

Voluntary participation in health-related studies is rarely 100%, and there is evidence that in recent years it has been declining (Hartge, 1999). A particular problem for epidemiological research arises when participation is not randomly distributed across study groups. Unfortunately, whether or not an individual agrees to participate in a project is often associated with the health outcome and with the exposure(s) under investigation; invariably this leads to biased estimates of risk (Rothman and Greenland, 1998). The assessment of the impact of differential participation requires both of those who do not take part to be compared with those who do. In most studies, however, information on non-participants is often sparse or non-existent. Indeed, in some designs, such as those that employ random digit dialling, investigators are not even able to identify non-participants (Wacholder et al., 1992).

Previous attempts aimed at exploring this form of selection bias have relied mainly on re-approaching non-participants and asking a restricted number of questions (e.g. Holt et al., 1991; Madigan et al., 2000; Wrensch et al., 2000). However, even when the identities of non-participants are known, practical and ethical considerations often prohibit the use of such methods. The design of the United Kingdom Childhood Cancer Study (UKCCS) provided a valuable opportunity to investigate the potential impact of participation bias (UKCCS Investigators, 2000). The findings, which used small-area census data to investigate differences between participating and non-participating subjects, are reported here.

Keywords: childhood cancer; aetiology; census; participation; case-control

Correspondence: Dr GR Law; E-mail: gr.law@leeds.ac.uk

*A complete list of investigators is given in The United Kingdom Childhood Cancer Study: Objectives, Materials and methods. British Journal of Cancer 2000: 82(5): 1073 – 1102. See Appendix for Management Committee, Regional Investigators and Processing Group

Received 16 November 2001; revised 29 November 2001; accepted 29 November 2001
same (former) Family Health Services Authority (England and Wales) or Health Board (Scotland) as their matched case. The first two controls selected were assigned as ‘first-choice’, following successful eligibility checks. In order to restrict the analysis to examining the effects of subject participation and to those with a valid address that could be linked to census, GP permission was obtained for all control families classified as ‘first-choice’. A proportion of the parents of first-choice controls declined to participate and the selection process was repeated until two control families were enrolled and interviewed. The two controls enrolled into the study were assigned as ‘interviewed’, with a proportion of these also assigned as first-choice.

**Linkage to the census of Great Britain**

Cases and controls were linked, via the postcode of their residence, to the 1991 census of GB, which was conducted around the time of the commencement of the study (The 1991 Census, Crown copyright, ESRC Purchase). Every household has a legal requirement to complete the census form, which contained 25 questions, the content ranging from age and sex of occupants to indicators of affluence such as adult employment and housing conditions. To maintain confidentiality, data are released at an areal level, with the smallest geographical units being indivisible and considered to be homogenous. The smallest units in England and Wales were the 108,336 Enumeration Districts (ED) and in Scotland the 38,084 Output Areas (OA). These can be aggregated into 9527 Electoral Wards (England and Wales) and 1002 Postcode Sectors (Scotland).

The address of the child at the time of diagnosis was used to link the study subjects to the census data. The address was assigned a validated postcode, using an automated system (QuickAddress™ and the Postal Address Books available from Royal Mail. All postcodes were assigned to an ED/OA using a lookup program PC2ED for England and Wales and the Postcode to OA databases for Scotland.

**Census data**

The census variables of interest, detailed in Table 1, focus on material possessions, employment, household characteristics, and population density. The simple counts provided by the census are given as a proportion of the relevant population (of households or people), as the geographically delimited census regions are not the same size in terms of population or spatial area. The index of deprivation and the assessment of population density are described in more detail below. From Table 2 onwards, all variables were divided into seven categories, with wherever possible an equal number of census units for GB in each group. An arithmetic mean for each census measure was calculated by weighting the GB data by the age, sex and regional distribution of cases from the study. The proportions of persons in each of the seven categories from the census, also weighted by the study values were calculated.

**Deprivation** An index of deprivation was generated for each ED/OA in GB (UKCCS Investigators, 2000). The proportions of unemployed economically active persons (aged 16 and over), households without a car and households not owner-occupied were calculated. Each proportion was transformed to give zero skew with a mean of zero and an overall standard deviation of one. The transformed value for each census unit was calculated and is equivalent to the number of standard deviations from the overall mean. These standard deviations, from the three variables, were summed for each unit and the resulting value is termed the index of deprivation. A high value positive value represents a region with high material deprivation and a high negative value represents an affluent area.

**Population density** The measures of population density are reported at the electoral ward/postcode sector level. The traditional estimate of population density \(d_i\) for each geographical region \(j\) is given as

\[
d_j = \frac{p_j}{a_j}
\]

where \(p_j\) is the number of persons and \(a_j\) the spatial area. This will be referred to as the ‘area-based’ population density and is expressed as persons per hectare. However, this does not necessarily reflect the density at which the ‘average’ person lives, but more closely reflects the land use and type of region. A more useful measure for the density at which an average person lives is termed the ‘person-weighted’ population density (Dorling and Atkins, 1995). This is defined as

\[
w_j = \sum_{i=1}^{n} d_i \left( \frac{p_i}{p_j} \right)
\]

where region \(j\) is divided into \(n\) smaller regions \(i\), \(d_i\) is the area-based population density and \(p_i\) the number of persons for the smaller region. For the purposes of this analysis, this may be regarded as a population-weighted sum of the area-based population densities for the ED/OA in each electoral ward/postcode sector. In order to provide a more intuitive assessment of urban-

**Table 1** Description and summary of the census variables used in the analysis

| Variable | Areal level* | GB weighted mean* | Min | Max |
|----------|--------------|-------------------|-----|-----|
| Index of deprivation | Composite index of unemployment, car and home ownership (see text) | ED/OA | 0.17 | -6.15 | 7.75 |
| Components of the deprivation index | | | | |
| Unemployed persons | Proportion economically-active persons unemployed | ED/OA | 0.11 | 0.00 | 1.00 |
| Houses with no car | Proportion of households without a car | ED/OA | 0.33 | 0.00 | 1.00 |
| Non-owner occupied houses | Proportion of households not owner-occupied | ED/OA | 0.35 | 0.00 | 1.00 |
| Other census variables | | | | |
| Household overcrowding | Proportion households with more than one person per room | ED/OA | 0.03 | 0.00 | 0.55 |
| Households with child dependants | Proportion households with at least one dependant child | ED/OA | 0.29 | 0.00 | 0.95 |
| Flat dwellings | Proportion dwellings which are flats | ED/OA | 0.18 | 0.00 | 1.00 |
| Households with central heating | Proportion households with central heating | ED/OA | 0.83 | 0.00 | 1.00 |
| Persons working over 40 hours/week | Proportion employed persons working over 40 h per week | ED/OA | 0.16 | 0.00 | 1.00 |
| Population density\(^{**}\) | Area-based population density (see text) | W/PC | 28.09 | 0 | 71.26 |
| Person-weighted | Person-weighted population density (see text) | W/PC | 40.37 | 0 | 998.68 |

*ED/OA-Enumeration district/output area; W/PC-ward/postcode area; \(^{**}\)Weighted for case distribution for sex, 5-year age group (0–14 year-olds) and study region; Based on persons per unit area.
rural status both measures of population density for all persons were categorized as urban (more than 25 persons per hectare), rural (less than 1.5 persons per hectare) or suburban (in between).

Statistical methods

The risk associated with each explanatory exposure measure is presented as odds ratios from a logistic regression model. Adjustment was made for single year of age, sex and study region as a representation of the matching variables (UKCCS Investigators, 1999). Two comparison groups, not mutually exclusive, were used: the first-choice controls and the interviewed controls.

RESULTS

Of the 4433 cases identified as eligible, permission to approach 4306 (97%) was obtained from the treating consultant, and 3838 (87%) agreed to participate (Table 2). Case interview rates varied by diagnostic group, from 93% for acute lymphoblastic leukaemia to 82% for the CNS tumours. Of the 7669 matched controls eligible for the study, 5530 (72%) agreed to participate; the proportion for control interview was more similar to the first-choice controls than to the interviewed controls. This resulted in the interviewed cases being from the most deprived seventh of GB, with a slightly lower proportion of controls from the most deprived group. The non-interviewed first-choice controls show a skew towards the more deprived groups, with 25% of those refusing to be interviewed coming from the most deprived seventh (where 15% would be expected). This resulted in the interviewed control group having fewer controls from the more deprived areas, with corresponding excesses in the more affluent areas. The distribution for interviewed cases was more similar to the first-choice controls, and to GB as a whole, than to the interviewed controls.

Table 3 shows the distribution of interviewed cases and control groups in the UK Childhood Cancer Study by deprivation category.

| Deprivation category | Interviewed cases | First-choice controls | GB % | Interviewed controls | GB % |
|----------------------|-------------------|------------------------|------|----------------------|------|
| Least                | 1041 (14)         | 183 (9)                | 2.3  | 1154 (15)            | 13   |
| Least                | 1090 (14)         | 204 (10)               | 2.3  | 1215 (16)            | 14   |
| Least                | 1057 (14)         | 221 (10)               | 2.3  | 1160 (15)            | 14   |
| Least                | 1109 (14)         | 260 (12)               | 2.3  | 1154 (15)            | 14   |
| Most                 | 1059 (14)         | 307 (14)               | 2.3  | 1038 (14)            | 14   |
| Most                 | 1097 (14)         | 384 (18)               | 2.3  | 990 (13)             | 14   |
| Most                 | 1162 (15)         | 526 (25)               | 2.3  | 918 (12)             | 17   |
| Total                | 3838 (100)        | 7669 (100)             |      | 2139 (100)           | 100  |

*Deprivation for the enumeration district/output Area; bWeighted by case distribution for a 5-year age group, study region and sex; cAddress not available to link census data.

DISCUSSION

Issues of participation are significant for all studies of disease aetiology that rely on individual compliance. The issue of non-participation, and consequential bias that may be introduced, is critically important in case–control studies that rely on personal contact to assess environmental experiences and exposures. Our findings indicate that in such studies, the profiling of non-participants may be as important as that of participants. In the study described here, the involvement of parents required active participation, as information was primarily collected through a face-to-face interview (UKCCS Investigators, 2000). As with all interview-based case-control studies, however, whilst the motivation of affected families to participate was strong, the motivation of those who were unaffected is less clear.

Comparisons of census data for cases and first-choice controls provided results that were free from participation bias: the findings suggesting that, on average, case families tended to live in areas that were more affluent than those of control families. In contrast, comparison with participating control families suggested that, on average, case families tended to live in areas that were less affluent than those of control families. This observation is consistent with other reports suggesting that participants often belong to a higher socio-economic groups than non-participants as measured by housing tenure, income, level of education and occupation (Holt...
et al., 1991; Hatch et al., 1998; Madigan et al., 2000; Wrensch et al., 2000).

Measures of material deprivation are often closely associated with possible aetiological factors such as smoking, occupation, and previous illness history. Indeed, in many epidemiological studies it is virtually impossible to identify potentially harmful exposures that are not—either directly or indirectly—related to measures of social class, deprivation or affluence. Rather, somewhat surprisingly, the treatment of bias as a real confounder (risk factor) rarely impacts on the magnitude of risk estimates (e.g. UKCCS Investigators, 2000).

### Table 3

| Deprivation | Cases | First-choice controls | Interviewed controls |
|-------------|-------|-----------------------|----------------------|
| GB%a        | n (%) | OR (95% CI)*          | n (%) | OR (95% CI)* |
| Least       | 13    | 512 (14)              | 1041 (14)          | 1154 (15) | 1154 (15) |
|              | 14    | 599 (16)              | 1090 (14)          | 1215 (16) | 1215 (16) |
|              | 14    | 591 (15)              | 1057 (14)          | 1160 (15) | 1160 (15) |
|              | 14    | 558 (15)              | 1109 (14)          | 1154 (15) | 1154 (15) |
|              | 14    | 504 (13)              | 1059 (14)          | 1038 (14) | 1038 (14) |
|              | 14    | 510 (13)              | 1097 (14)          | 990 (13)  | 990 (13)  |
| Most        | 17    | 544 (14)              | 1162 (15)          | 918 (12)  | 918 (12)  |

Component of the deprivation index

### Unemployment

| Least | 12 | 503 (13) | 910 (12) | 1041 (14) | 1154 (15) |
|-------|----|----------|----------|-----------|-----------|
|       | 15 | 593 (15) | 1192 (16) | 1297 (17) | 1297 (17) |
|       | 14 | 580 (15) | 1158 (15) | 1247 (16) | 1247 (16) |
|       | 14 | 564 (15) | 1041 (14) | 1094 (14) | 1094 (14) |
|       | 14 | 531 (14) | 1056 (14) | 1043 (14) | 1043 (14) |
|       | 15 | 518 (14) | 1080 (14) | 956 (13)  | 956 (13)  |
|       | 17 | 549 (14) | 1178 (15) | 951 (12)  | 951 (12)  |

### Car ownership

| Least | 12 | 424 (11) | 863 (11) | 714 (9)  | 120 (14)  |
|-------|----|----------|----------|----------|-----------|
|       | 15 | 610 (16) | 1120 (15) | 1223 (16) | 1223 (16) |
|       | 14 | 595 (16) | 1158 (15) | 1274 (16) | 1274 (16) |
|       | 14 | 570 (15) | 1115 (15) | 1157 (15) | 1157 (15) |
|       | 14 | 551 (14) | 1145 (15) | 1057 (14) | 1057 (14) |
|       | 15 | 472 (12) | 1041 (14) | 932 (12)  | 932 (12)  |
|       | 17 | 549 (14) | 1178 (15) | 951 (12)  | 951 (12)  |

### Home ownership

| Least | 12 | 611 (17) | 1230 (16) | 1413 (17) | 1413 (17) |
|-------|----|----------|----------|----------|----------|
|       | 15 | 659 (17) | 1223 (16) | 1327 (17) | 1327 (17) |
|       | 14 | 546 (14) | 1111 (14) | 1135 (15) | 1135 (15) |
|       | 14 | 535 (14) | 1053 (14) | 1052 (14) | 1052 (14) |
|       | 14 | 511 (13) | 1007 (13) | 995 (13)  | 995 (13)  |
|       | 14 | 513 (13) | 1009 (13) | 989 (12)  | 989 (12)  |
|       | 13 | 413 (11) | 982 (11)  | 809 (11)  | 809 (11)  |

### Other measures

#### Household overcrowding

| Least | 18 | 788 (21) | 1463 (19) | 1587 (21) | 1587 (21) |
|-------|----|----------|----------|----------|----------|
|       | 15 | 629 (16) | 1114 (15) | 1242 (16) | 1242 (16) |
|       | 14 | 552 (14) | 1094 (14) | 1118 (15) | 1118 (15) |
|       | 13 | 485 (13) | 991 (13)  | 989 (13)  | 989 (13)  |
|       | 13 | 491 (13) | 1003 (13) | 980 (13)  | 980 (13)  |
|       | 13 | 454 (12) | 993 (13)  | 956 (13)  | 956 (13)  |
|       | 13 | 439 (11) | 957 (12)  | 757 (10)  | 757 (10)  |

#### Households with child dependants

| Least | 13 | 419 (11) | 957 (12)  | 757 (10)  | 757 (10)  |

#### Flat dwellings

| Least | 21 | 851 (22) | 1595 (21) | 1667 (22) | 1667 (22) |
|-------|----|----------|----------|----------|----------|
|       | 17 | 620 (16) | 1264 (16) | 1313 (17) | 1313 (17) |
|       | 15 | 588 (15) | 1179 (15) | 1213 (16) | 1213 (16) |
|       | 16 | 575 (15) | 1139 (15) | 1154 (15) | 1154 (15) |
|       | 14 | 550 (14) | 1093 (14) | 1062 (14) | 1062 (14) |
|       | 11 | 497 (12) | 838 (11)  | 793 (10)  | 793 (10)  |
|       | 7  | 236 (6)  | 507 (7)   | 427 (6)   | 427 (6)   |

Continued
The challenge is to disentangle the artifactual consequences of participation bias from genuine aetiological factors. Some researchers have suggested using the variable most closely related to participation as a confounder to ‘adjust’ for participation. Such adjustments are, however, only appropriate when the antecedents of both exposure and disease, or their joint distributions, are available for the entire study population (Greenland, 1998). As in the majority of studies, neither of these were available for the UKCCS, where individual reasons for participation could only be indirectly related with the deprivation index.

As expected, our findings show that the primary care sampling frame provided control families that were broadly representative of the general GB population, where the majority of people register with a GP in the area in which they live – access to NHS medical care requiring registration which covers approximately 98% of the population (RCGP, 1987). The UKCCS sampling frame compares favourably with other methods of control selection. For example, random digit dialling, a method commonly employed in the USA and Canada (Robison and Daigle, 1984), prohibits collection from homes without a telephone, those not at home when telephoned, and those who answer but who either refuse to answer any questions or deliberately lie about their families eligibility.

In conclusion, our findings confirm that differential participation is a potentially major source of bias in case–control studies that estimate risks on the basis of information reported from respondents alone. Studies that ignore this source of bias may produce misleading results. There is a clear need to address this issue in terms of study design, and in the application of appropriate statistical methods to try to overcome this bias.

ACKNOWLEDGEMENTS

The UKCCS was sponsored and administered by the United Kingdom Co-ordinating Committee on Cancer Research and was supported by the UKCCS paediatric oncologists and by the National Radiological Protection Board. Financial support was provided by: Cancer Research Campaign, Imperial Cancer Research Fund, Leukaemia Research Fund, and Medical Research Council through grants to their units; Leukaemia Research Fund, Department of Health, member companies of Electricity Association, Irish Electricity Supply Board, National Grid Company plc, and Westlakes Research (Trading) Ltd for general expenses of the study; Kay Kendall Leukaemia Fund for associated laboratories studies; and Foundation of Children with Leukaemia for study of electrical fields. The investigation in Scotland is funded by The Scottish Office, Scottish Power plc, Scottish Hydro-Electric plc, and Scottish Nuclear Ltd. We thank the members of the UKCCSG for their support, the staff of the local hospitals, the family physicians and their practice staff. We especially thank the families of the children included in the study for their help.
REFERENCES

Dorling D, Atkins D (1995) Population density, change and concentration in Great Britain 1971, 1981 and 1991. OPCS Series SMPS no 58. London: HMSO

Greenland S (1998) Basic methods for sensitivity analysis and external adjustment. In Modern Epidemiology, 2nd edn, Rothman KJ, Greenland S (eds) pp 115 – 134. Philadelphia: Lippincott-Raven

Hartge P (1999) Raising response rates: Getting to yes. Epidemiology 10: 105 – 107

Hatch EE, Linet MS, Kleinerman RA, Tarone RE, Severson RK, Hartsock CT, Haines C, Kaune WT, Friedman D, Robison LL, Wacholder S (1998) Association between childhood acute lymphoblastic leukaemia and use of electrical appliances during pregnancy and childhood. Epidemiology 9: 234 – 245

Holt VL, Daling JR, Stergachis A, Voigt LF, Weiss NS (1991) Results and effect of refusal recontact in a case-control study of ectopic pregnancy. Epidemiology 2: 375 – 379

Madigan MP, Troisi R, Potischman N, Brogan D, Gammon MD, Malone KE, Brinton LA (2000) Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women. International J Epidemiol 29: 793 – 798

APPENDIX 1

Writing Committee – GR Law, AG Smith, E Roman

Management Committee – KK Cheng, Central Region; N Day, East Anglia Region; R Cartwright, A Craft, North East Region; JM Birch, OB Eden, North West Region; PA McKinney, Scotland; J Peto, South East Region; V Beral, E Roman, South Midlands Region; P Elwood, South Wales Region; FE Alexander, South West Region; CED Chilvers, Trent Region; R Doll, Epidemiological Studies Unit, University of Oxford; GM Taylor, Immunogenetics Laboratory, University of Manchester, Manchester; M Greaves, Leukaemia Research Fund Centre, Institute of Cancer Research; D Goodhead, Radiation and Genome Stability Unit, Medical Research Council, Harwell; FA Fry, National Radiological Protection Board; G Adams, UK Co-ordinating Committee for Cancer Research.

Regional Investigators – KK Cheng, E Gilman, Central Region; N Day, J Skinner, D Williams, East Anglia Region; R Cartwright, A Craft, North East Region; JM Birch, OB Eden, North West Region; PA McKinney, Scotland; J Deacon, J Peto, South East Region; V Beral, E Roman, South Midlands Region; P Elwood, South Wales Region; FE Alexander, M Mott, South West Region; CED Chilvers, K Muir, Trent Region.

Data Processing Group – GR Law, J Simpson, E Roman.

A complete list of investigators is given in: The United Kingdom Childhood Cancer Study: objectives, materials, and methods. Br J Cancer 2000; 82: 1073 – 1102.