

## **Who are we missing? Area deprivation and survey participation**

Anna Goodman<sup>1\*</sup> and Rebecca Gatward<sup>2</sup>

<sup>1</sup> London School of Hygiene and Tropical Medicine, London, UK

<sup>2</sup> Office for National Statistics, London, UK

**Corresponding author:** Anna Goodman, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK ([anna.goodman@lshtm.ac.uk](mailto:anna.goodman@lshtm.ac.uk)).

*Note: this is a personal version, created by Anna Goodman, of the text of the accepted journal article. It reflects all changes made in the peer review process, but does not incorporate any minor modifications made at the proof stage. The complete citation for the final journal article is:*

- Goodman, A; Gatward, R; (2008) Who are we missing? Area deprivation and survey participation. *Eur J Epidemiol*, 23 (6). pp. 379-87.
- DOI: 10.1007/s10654-008-9248-0

Copyright © and Moral Rights for this paper are retained by the individual authors and/or other copyright owners

## **Abstract**

This study investigates selective participation by area deprivation in two population-based, nationally-representative surveys of children aged 5-16 years in England. We analysed the English participants (N=22 903) in the British Child and Adolescent Mental Health Surveys, conducted in 1999 and 2004. The surveys recruited parents, teachers and children, allowing us to calculate participation rates for all three types of informant, with particularly detailed analyses possible for parents. Area-deprivation was assigned using Indices of Multiple Deprivation, 2004. We found that parents, teachers and children from more deprived areas were all substantially less likely to participate, with participation 8-9% lower in the most deprived areas than in the least deprived areas. There was substantial heterogeneity in the effect of deprivation on different types of parental participation, with a strong association for overall non-contact but none for overall non-cooperation. Refusal to participate in response to opt-out letters showed the opposite gradient to all other forms of non-participation, with individuals from less deprived areas being more likely to opt out. These findings indicate that the deprivation level of area of residence predicts non-response in multiple types of informants, which may exacerbate bias when estimating the prevalence of health outcomes. Furthermore, the magnitude and the direction of this may depend on the recruitment method used. Differential response by area deprivation may therefore be worth measuring in any survey.

**Keywords:** Area deprivation, Response rates, Survey participation, Selective participation

**Abbreviations:** **B-CAMHS** = British Child and Adolescent Mental Health Surveys; **CBR** = Child Benefit Register; **IMD** = Index of Multiple Deprivation; **ONS** = Office for National Statistics; **SDQ** = Strengths and Difficulties Questionnaire.

## **Introduction**

When individuals who take part in epidemiological studies differ systematically from those who do not, this can threaten generalisability and external validity. Such selective participation has been demonstrated in relation to individual and area-level factors in many high income countries (1-3), and may be exacerbated by the apparent decline in survey participation in the UK and the US in recent years (4, 5). Knowing the magnitude, direction and explanation for selective participation is therefore important for drawing valid inferences and for focussing efforts on reducing such biases in future research. Yet this is hard to do when little is typically known about who is missing from our surveys.

In surveys using address-based registers as a sampling frame, postcode of residence is often available for respondents and non-respondents alike; although access restrictions may apply in publicly available datasets. This opens the possibility of examining response rates (strictly ‘proportions’) according to postcode-derived ecological characteristics such as area deprivation. Several large population-based epidemiological surveys in the UK have used this strategy in recent years, looking at health and non-health outcomes in a range of study designs (1, 6-10). All eleven examples found strong evidence of lower response rates in more deprived areas, while to our knowledge no studies have found the reverse. Seven of the eleven surveys further suggest

that the effect of deprivation may be greater upon participant non-contact than upon non-participant non-cooperation (1, 7, 8), although only one tests this formally (8).

While presenting a fairly consistent picture, these studies have several important limitations. Several use deprivation measures based on relatively few area-level characteristics collected up to a decade previously (1, 8-11). Others simply dichotomise areas into ‘advantaged’ and ‘disadvantaged’ wards (6, 7), or look for heterogeneity within the category of ‘deprived’ areas (10); only one study provides sufficient raw data to examine whether area deprivation has a dose response effect on response rates across the whole spectrum of deprivation (9). As such, it is unclear whether it is only in the most deprived areas that non-response is lower. This is of interest because, if so, the problem of deprivation-related response bias could potentially be addressed by targeting extra resources into increasing response in highly deprived areas.

Classification of reasons for non-participation has been similarly crude. Most studies simply report non-participation, and no study goes beyond a binary categorisation of non-participation (usually non-contact vs. non-cooperation, although one study compares consenting to receive a questionnaire with subsequently completing that questionnaire (10)). Yet without knowing *why* people are not taking part, it is very hard to know best how to increase response rates; the best strategy for dealing with non-response because of out-of-date postal addresses will not be the same as the best strategy for dealing with non-response because potential participants do not feel motivated to take part.

Finally, while these examples include studies with a focus on both adults (1, 8, 10) and children (6, 7, 9, 11), the surveys of children sought only to interview the children’s parents, and not the children themselves. To our knowledge, no investigation has investigated child participation, or surveys in which some informants act as ‘gatekeepers’ determining whether other informants may be approached. Such gatekeeper designs are very common in surveys of children, for which active parental consent is increasingly a requirement (12), but there are also some examples for research with adults, such as in surveys using proxy informants (for example, 13).

In this paper we address some of these limitations using data from two nationally-representative, population-based surveys in England.

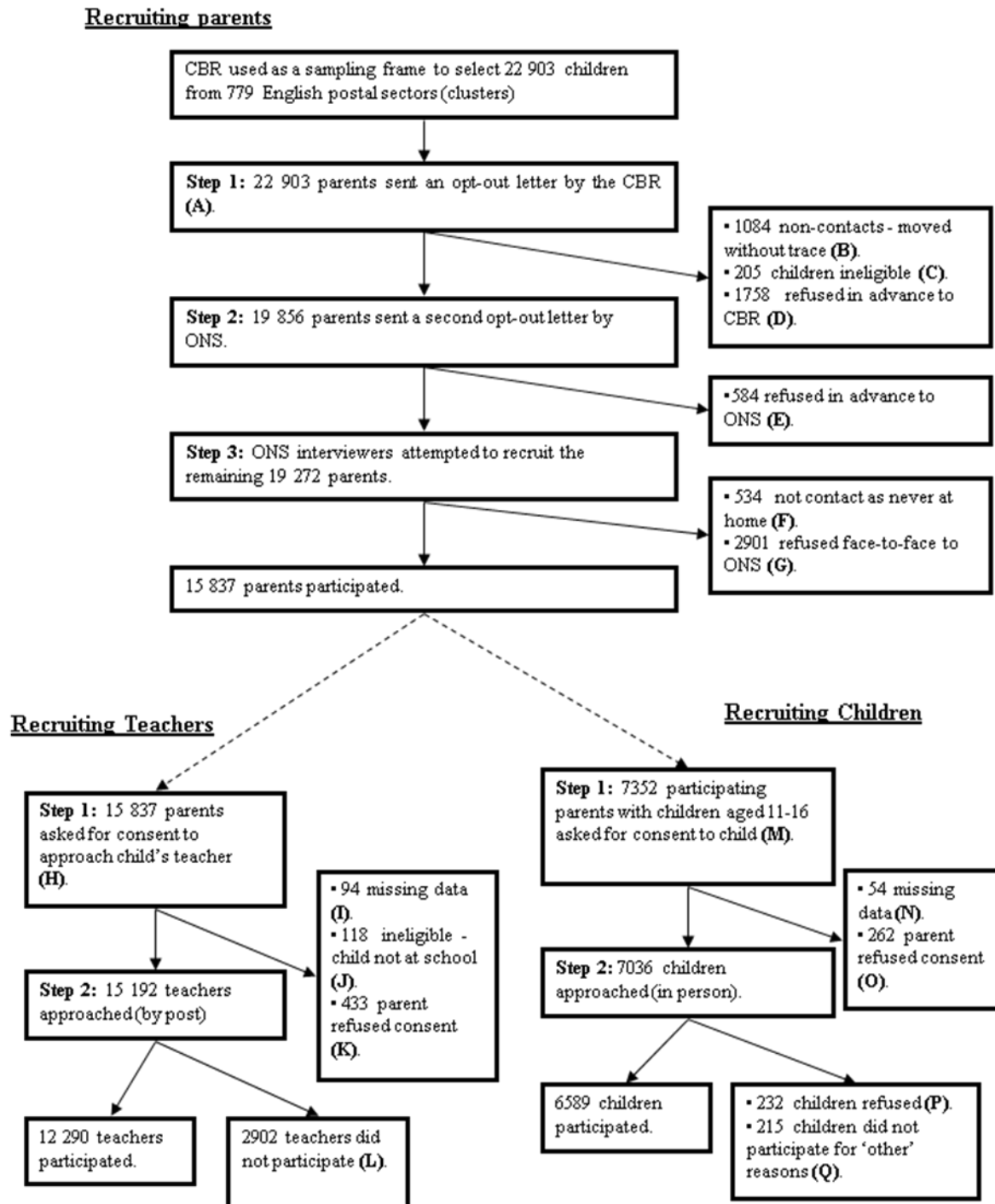
## **Methods**

### **Survey data and recruitment methodology**

This paper uses the English subsamples of the 1999 and 2004 British Child and Adolescent Mental Health Surveys (B-CAMHS), full details of which have been reported previously (14-16). Briefly, B-CAMHS were two independent cross-sectional surveys of the mental health of children and adolescents in Great Britain, carried out by the Office for National Statistics (ONS), and sampling children aged 5-15 in 1999 and 5-16 in 2004. Both surveys used the British Child Benefit Register (CBR) as a nationally-representative sampling frame, employing a clustered design by postal sector. The principal caregivers (‘parents’) of selected children were approached to take part in the B-CAMHS surveys. Conditional upon parental participation and consent, the teacher nominated by the parent as ‘knowing the child best’ was approached to

participate, as were children aged 11-16. This nested recruitment process is outlined in Figure 1. Ethical approval for the B-CAMHS studies was granted by the Institute of Psychiatry, London.

Figure 1: Recruitment of parents, teachers and children into the B-CAMHS surveys



CBR = Child Benefit Record, ONS = Office for National Statistics

## Defining participation and participation rates

Parents, teachers and children were defined as participating if they completed all or part of the brief mental health measure the Strengths and Difficulties Questionnaire (SDQ; 17). In addition to total non-participation, we defined *a priori* four sub-types of non-participation for parents; two for teachers; and three for children (see Tables 1, 2 and 3). All non-participation rates are calculated in accordance with standard definitions from the American Association for Public Opinion Research (18), at each stage taking as the denominator those individuals who could theoretically feature in the numerator (for details, see key to Tables 1, 2 and 3).

## Measuring small-area deprivation

The Index of Multiple Deprivation 2004 (IMD 2004) is a small-area measure of neighbourhood deprivation, calculated as the weighted average across seven domains of deprivation using data from the years 2000 to 2003 (19). The seven domains of deprivation are income; employment; health and disability; education, skills and training; barriers to housing and services; crime; and living environment. IMD scores are calculated at the level of Lower Super Output Areas (LSOAs), these being geographical units of varying size but containing a population of approximately 1500 at the time of the 2001 census (20).

20 050 (or 87.5%) of the 22 903 children originally selected by the CBR as potential participants were assigned IMD 2004 scores using their current postcode. For two sorts of non-participant, however, current postcode was not known and IMD score therefore had to be imputed. The families of 4.7% (N= 1085) of the study children had moved without trace (see Figure 1) and their current address was, by definition, unknown. For these children, we assigned the IMD score corresponding with the postcode of their last known address. For a further 7.6% of study children (N = 1741) we had no current postcode because the child’s parent refused in advance to participate when the CBR first sent an opt-out letter, and the ONS was therefore never given their address. For these individuals we imputed IMD decile on the basis of their postcode sector (i.e. the sampling cluster). We also imputed IMD decile in this way for the 28 additional individuals with postcodes which were missing or otherwise invalid. Postcode sectors contain an average of 6000 individuals in 2550 households.

## Statistical methods

Participation results are analysed by IMD decile, as defined using observed IMD in our sample. For logistic regression, IMD decile was treated as a continuous variable with all models being assessed for whether a quadratic term improved model fit – unless stated otherwise, . All analyses adjust for the clustered sampling design and present the pooled results across the 1999 and the 2004 B-CAMHS surveys; the pattern of the findings is unchanged when the surveys are analysed separately. For children whose IMD score was missing because they had no valid postcode, we performed multiple imputation using an approximate Bayesian Bootstrap (21). This model assumes that within a given postcode sector individuals were missing completely at random with respect to their IMD decile. For a more detailed description of our imputation method, see online supplementary material [Supplementary Material 1]. We performed most analyses in STATA 9.2. The exceptions were using Sas v.9.1 to fit multilevel random effects models and using Mplus 3 to generate a latent measure of individual socio-economic position.

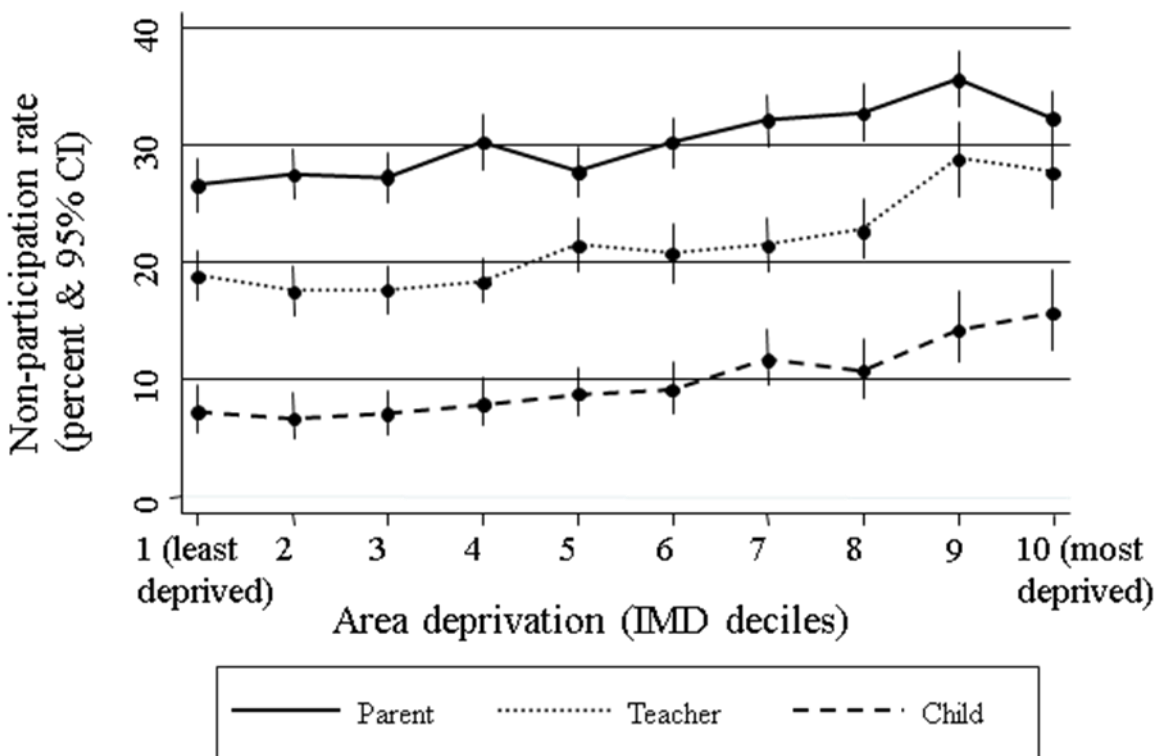
## Results

### Overall non-response by area deprivation decile

Between the two B-CAMHS surveys, 22 903 study children were initially selected from English postal sectors; the number of individuals at each subsequent stage is shown in Figure 1. The mean age of these children was 10.2 years and 49% were female, these figures being similar across IMD deciles (mean age 10.0-10.3 years; proportion female 47%-51%). We do not adjust for age and sex in subsequent regression analyses, because these are not known for children whose parents refused in advance to the CBR. Where age and sex were known for all informants (two parent variables and all teacher and child variables), adjusting for age and sex left the point estimates and confidence intervals for the effect of IMD decile unchanged to 2 decimal places.

Overall non-response rates were 30% in parents, 21% in teachers and 10% in children. There is strong evidence ( $p < 0.001$ ) that deprivation predicts non-participation in parents, teachers and children, all of whom show higher non-participation with increasing deprivation (see Figure 2, and the first columns of Table 1, 2 and 3). For most non-participation variables, deprivation appears to have a linear effect across the whole range, although in three cases a quadratic term is also predictive.

Figure 2: Parent, teacher and child non-response by IMD decile



### Types of non-participation for parents

There is strong evidence that the gradient of the effect of area deprivation differs between overall parent non-contact and overall parent non-cooperation ( $p < 0.001$ ). Parent non-contact, like overall parent non-response, shows strong evidence of a relatively large gradient with area deprivation (OR 1.11 per increase in IMD decile, 95% CI 1.09 - 1.14,  $p < 0.001$ ). By contrast, the association with parental non-cooperation is much smaller and of only borderline significance (OR 1.01, 95% CI 1.00 - 1.03,  $p = 0.04$ ).

Table 1 goes beyond this non-contact/non-cooperation dichotomy, however, and presents the four participation sub-types determined *a priori*. It is striking that while all four show strong evidence of a participation gradient by IMD decile, the *direction* of the effect differs. Non-contact due to the family moving without trace or because parents were never at home are both more common in more deprived areas, showing a large and progressive increase across the whole range. Face-to-face non-cooperation (due in 97% of cases to parent refusal) also shows a progressive, although less steep, increase in more deprived areas.

**Table 1: Parent non-participation - by area deprivation**

	Parent non-response among those eligible	Parent non-contact as moved without trace	Parent non-contact as never home	Parent refusal in advance	Parent face-to-face non-cooperation
<b>Overall non-participation</b>					
N	6 861/ 22 698	1 084/22 903	534/19 272	2 342/19 272	2 899/18 728
% and 95% CI	30.2 (29.5 – 31.0)	4.27 (4.43 – 5.06)	2.77 (2.43 – 3.16)	10.8 (10.4 – 11.3)	15.5 (14.8 – 16.1)
<b>Non-participation by IMD decile (% and 95% CI)</b>					
1 (least deprived)	26.5 (24.3 - 28.7)	2.8 (2.1 - 3.6)	1.3 (0.8 - 2.0)	12.8 (11.2 - 14.4)	12.1 (10.5 - 13.8)
2	27.7 (25.5 - 29.8)	3.4 (2.6 - 4.2)	1.1 (0.7 - 1.7)	12.1 (10.3 - 13.9)	13.7 (11.9 - 15.7)
3	27.3 (25.1 - 29.4)	4.0 (3.1 - 4.9)	1.7 (1.2 - 2.4)	11.4 (9.7 - 13.2)	12.8 (11.2 - 14.7)
4	30.4 (28.0 - 32.7)	4.6 (3.7 - 5.6)	2.8 (2.1 - 3.7)	11.5 (10.0 - 13)	15.0 (13.3 - 16.9)
5	27.8 (25.6 - 30.0)	4.2 (3.3 - 5.2)	1.7 (1.2 - 2.5)	9.8 (8.0 - 11.6)	14.9 (13.2 - 16.8)
6	30.1 (27.8 - 32.4)	5.0 (4.0 - 5.9)	3.0 (2.2 - 4.0)	10.9 (9.1 - 12.7)	14.9 (13.3 - 16.6)
7	32.1 (29.9 - 34.4)	5.3 (4.3 - 6.3)	3.4 (2.6 - 4.6)	11.0 (9.4 - 12.6)	16.7 (14.9 - 18.7)
8	32.8 (30.4 - 35.1)	6.5 (5.3 - 7.6)	3.1 (2.2 - 4.2)	10.6 (8.9 - 12.3)	17.2 (15.1 - 19.4)
9	35.5 (33.2 - 37.9)	5.9 (5.0 - 6.9)	5.0 (3.8 - 6.4)	9.1 (7.6 - 10.6)	20.7 (18.8 - 22.8)
10 (most deprived)	32.3 (30.0 - 34.7)	5.7 (4.6 - 6.8)	4.7 (3.3 - 6.5)	8.9 (7.5 - 10.3)	17.2 (15.3 - 19.4)
<b>OR &amp; 95% CI per increase in IMD decile</b>					
Linear term	1.04 (1.03 - 1.05)***	1.08 (1.05 - 1.1)***	1.17 (1.12 - 1.22)***	0.96 (0.94 - 0.98)***	1.06 (1.04 - 1.08)***
Quadratic term	--	--	--	--	--

\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ . Note that different denominators mean results in any row add to more than 100% **Key to non-response categories in Table 1 (with reference to Figure 1):** Non-response among those eligible =  $(B + D + E + F + G)/(A - C)$ ; Parent non-contact as moved without trace =  $B/A$ ; Parent non-contact as never home =  $F/(A - B - C - D - E)$ ; Refusal in advance;  $(D+E)/(A - B - C)$ ; Refusal face-to-face =  $G/(A - B - C - D - E - F)$ .

By contrast, non-participation due to refusal in advance *decreases* with increasing deprivation. The difference is not very large; the rate of refusal in advance is 12.8% in the least deprived decile vs. 8.9% in the most deprived decile. The difference is, however, internally replicated if one looks separately at the two points at which parents were given the opportunity to refuse in advance; first to the CBR and subsequently to the ONS. In both cases there is strong evidence ( $p < 0.001$ ) of a progressive decrease in refusal in advance with increasing area deprivation (OR for refusal in advance to CBR 0.97, 95% CI 0.95-0.99,  $p = 0.004$ ; OR for refusal in advance to ONS 0.95, 95% CI 0.92 – 0.98,  $p = 0.001$ ).

### Types of non-participation for teachers and children

For total teacher non-response there was evidence that a quadratic term improved model fit, with the positive value of the quadratic indicating that the effect of deprivation was particularly large in more deprived areas. This non-linear relationship was driven by ‘parent refused consent to approach teacher’ which only increased notably in the most deprived deciles and indeed shows some suggestion of a U-shaped relationship, although the absolute differences are small. By contrast, teacher non-response – which was much the more common reason for teacher non-participation – shows a linear effect of increasing deprivation predicting higher non-response across the range (see Table 2).

**Table 2: Teacher non-participation, by area deprivation**

	Teacher non-response among those eligible	Parent refused consent to approach teacher	Teacher non-response when parent gave consent
<b>Overall non-participation</b>			
N	3 335 / 15 616	433 / 15 616	2 902 / 15 183
% and 95% CI	21.4 (20.45 - 22.3)	2.8 (2.5 - 3.1)	19.1 (18.3 - 20.0)
<b>Non-participation by IMD decile (% and 95% CI)</b>			
1 (least deprived)	18.9 (16.9 - 21.0)	3.4 (2.3 - 4.8)	16.0 (14.2 - 18.1)
2	17.5 (15.6 - 19.7)	2.7 (1.9 - 3.7)	15.3 (13.5 - 17.3)
3	17.6 (15.7 - 19.7)	2.4 (1.7 - 3.4)	15.6 (13.7 - 17.7)
4	18.3 (16.5 - 20.3)	2.5 (1.8 - 3.5)	16.2 (14.6 - 18.0)
5	21.5 (19.3 - 23.9)	2.7 (1.9 - 3.6)	19.3 (17.2 - 21.7)
6	20.7 (18.3 - 23.3)	2.3 (1.5 - 3.4)	18.8 (16.4 - 21.5)
7	21.5 (19.3 - 23.8)	1.9 (1.3 - 2.8)	19.9 (17.8 - 22.2)
8	22.7 (20.3 - 25.4)	1.9 (1.4 - 2.9)	21.2 (18.8 - 23.8)
9	28.8 (25.9 - 32.0)	4.0 (2.9 - 5.5)	25.8 (22.8 - 29.1)
10 (most deprived)	27.7 (24.7 - 31.0)	4.1 (2.9 - 5.7)	24.7 (21.9 - 27.7)
<b>OR &amp; 95% CI per increase in IMD decile</b>			
Linear term	0.97 (0.91 - 1.05)	0.76 (0.63 - 0.91)**	1.08 (1.06 - 1.10)***
Quadratic term	1.01 (1.00 - 1.02)**	1.03 (1.01 - 1.04)***	--

\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ . Note that different denominators mean results in any row add to more than 100%.

**Key to non-response categories in Table 2 (with reference to Figure 1):** Teacher non-response among those eligible (i.e. where the parent took part and the child was at school) =  $(K + L)/(H - I - J)$ ; Parent refused consent to approach teacher =  $K/(H - I - J)$ ; Teacher non-response when parents gave consent =  $L/(H - I - J - K)$ .



For child non-participation, the high non-response rates in more deprived areas result both from parents being more likely to refuse consent to allow a child to participate, and children who were eligible to be approached failing to take part (see Table 3). A quadratic term improved the fit of the ‘parent refused consent to approach child’, with refusal only increasing notably in the most deprived areas. In total 447 children did not take part despite being eligible to be approached: of these 232 (52%) refused to take part, 34 were ‘incapable’ (e.g. due to severe disability) (8%), 12 were at boarding school (3%) and 169 did not take part for ‘other’ reasons (38%). In the absence of more information about the ‘other’ reasons why children did not take part, there is insufficient information to calculate non-contact and non-cooperation rates. It is, however, possible to calculate child refusal rates; these do show a gradient by deprivation, but not strikingly more so than the combination of all reasons for child non-response (see columns 4 and 5 of Table 3).

**Table 3: Child non-participation, by area deprivation**

	Child non-response among those eligible	Parent refused consent to approach child	Child non-response when parent gave consent	Child refusal
<b>Overall non-participation</b>				
N	709 / 7 291	262/7 029	447/6 582	232/6 797
% and 95% CI	9.7 (8.9 - 10.6)	3.6 (3.1 - 4.2)	6.4 (5.7 - 7.17)	3.3 (2.8 - 3.8)
<b>Non-participation by IMD decile (% and 95% CI)</b>				
1 (least deprived)	7.3 (5.5 - 9.5)	2.9 (1.7 - 4.8)	4.5 (3.4 - 6.1)	2.6 (1.7 - 3.9)
2	6.7 (5.0 - 9.1)	2.9 (1.7 - 4.6)	4 (2.7 - 5.8)	2 (1.2 - 3.3)
3	7.1 (5.4 - 9.1)	2.2 (1.3 - 3.6)	5 (3.6 - 6.8)	2.7 (1.7 - 4.3)
4	7.9 (6.1 - 10.2)	2.3 (1.4 - 3.7)	5.8 (4.3 - 7.6)	2.2 (1.3 - 3.5)
5	8.8 (6.9 - 11.1)	3.5 (2.5 - 5.1)	5.5 (4.0 - 7.4)	2.9 (1.9 - 4.5)
6	9.1 (7.0 - 11.7)	2.6 (1.7 - 4.2)	6.6 (4.9 - 9.0)	4.3 (2.9 - 6.5)
7	11.7 (9.5 - 14.4)	4.0 (2.8 - 5.7)	8.1 (6.1 - 10.5)	2.9 (1.9 - 4.56)
8	10.7 (8.4 - 13.5)	3.1 (1.9 - 4.9)	7.9 (6.0 - 10.3)	3.7 (2.5 - 5.5)
9	14.2 (11.5 - 17.5)	6.2 (4.4 - 8.5)	8.6 (6.5 - 11.2)	4.9 (3.3 - 7.1)
10 (most deprived)	15.7 (12.5 - 19.5)	7.3 (5.3 - 10.0)	9 (6.7 - 12.1)	5.3 (3.6 - 7.8)
<b>OR &amp; 95% CI per increase in IMD decile</b>				
Linear term	1.11 (1.08 - 1.15)***	0.83 (0.67 - 1.03)	1.10 (1.06 - 1.14)***	1.10 (1.05 - 1.16)***
Quadratic term	--	1.03 (1.01 - 1.04)**	--	--

\* p≤0.05, \*\* p≤0.01, \*\*\* p≤0.001. Note that different denominators mean results in any row add to more than 100%.

**Key to non-response categories in Table 3 (with reference to Figure 1):** Child non-response among those eligible (i.e. where the parent took part and the child was 11-16) =  $(O + P + Q)/(M - N)$ ; Parent refused consent to approach child =  $O/(M - N)$ ; Child non-response when parents gave consent =  $(P + Q)/(M - N - O)$ ; Child refusal rate =  $(P)/(M - N - O)$ .

### Sensitivity analyses regarding the validity of the imputation model

As described above, the IMD decile of parents who refused in advance to the CBR were imputed under the assumption of ‘missing completely at random’ within any given postcode sector. We conducted sensitivity analyses assuming that those who refused in advance to the CBR were drawn systematically either from the most deprived or the least deprived Lower Super Output Areas (LSOAs) in their postcode sector. The overall effect of deprivation on parent non-response

was robust, but, the association with refusal in advance – and specifically refusal in advance to the CBR which relies upon imputed data - was sensitive even to small changes. Under the assumption that it was individuals from the more deprived LSOAs within the less deprived postal sectors who were missing, the direction of the association even reversed. There was, however, no evidence that an ecological fallacy of this sort was likely to apply at the level of the postcode sector in any of the other 11 parent, teacher and child non-participation variables listed in Tables 1, 2 or 3. This was demonstrated by multilevel random effects models (modelling LSOAs within postcode sectors) fitted only to the *observed* IMD deciles. In all cases a within-cluster regression coefficient of the same sign as the between-cluster coefficient. There was likewise no evidence of an ecological fallacy for refusal in advance to the ONS, this being the non-participation category most similar to refusal in advance to the CBR. For more detailed results, see online supplementary material [Supplementary material 2].

### **Sensitivity analyses in teachers and children of the effect using of small-area deprivation not individual-level indicators of socio-economic deprivation**

The above analyses explore the possibility of an ecological fallacy when moving from the postcode sector to the LSOAs. It is, however, also possible that there could be an ecological fallacy *within* LSOAs – for example, if the less deprived individuals within deprived LSOAs did not take part. We explored this for teachers and children using an individual-level measure of family socio-economic deprivation, created using parent-reported information on maternal education, household income, housing tenure and parental employment. The Pearson’s correlation between IMD decile and this individual-level measure of family socio-economic deprivation was 0.51. Repeating the logistic regression analyses reported in Tables 2 and 3 resulted in very similar substantive findings and a multilevel random effects model (modelling individuals within LSOAs) demonstrated in all cases a within-cluster regression coefficient of the same sign as the between-cluster coefficient. There is therefore no evidence of an ecological bias within LSOAs of a sort which would change the substantive conclusions of this paper. For more information about the creation of the latent variable and for detailed results, see online supplementary material [Supplementary material 3].

### **Mental health as a possible mediator of non-response in teachers and children**

Previous analyses of the B-CAMHS data demonstrates that children living in more socio-economically disadvantaged areas have higher rates of mental health problems (14, 15). To explore whether higher non-participation rates in more deprived areas might be the product of a higher rate of child mental health problems, we repeated the above analyses after additionally entering into the regression model the child’s score on the parent Strengths and Difficulties Questionnaire. In all seven cases this changed the estimated odds ratios and confidence intervals associated with deprivation very little (usually identical to two decimal places).

## **Discussion**

Our study of response rates in two population-based, cross-sectional surveys of child mental health in England provides strong evidence that parental non-response was progressively higher in more deprived areas. In most cases this effect appeared to be linear across the whole range, although in a few cases the increase appeared particularly pronounced in more deprived areas.

Within this overall pattern there was evidence of marked internal heterogeneity for different types of response. Non-contact, whether due to the family moving without trace or not being at home whenever the interviewer visited, was substantially more common in children whose last known address/current address was in a more deprived area. By contrast, overall parental non-cooperation showed only a small difference by area deprivation. One implication is that surveys such as B-CAMHS may underestimate the prevalence of health (or other) outcomes which are more common in more deprived areas. Another implication is that research teams may be able to reduce the deprivation gradient by investing more resources in making contact with potential participants, as these are the larger contributor to differential non-participation.

Thus far, therefore, our study replicates the work reviewed in the introduction. It also extends these findings with what is, to our knowledge, a novel research finding regarding heterogeneity in *when* parents in different areas declined to take part. Parents in less deprived areas were more likely to refuse in advance in response to opt-out letters, while parents in more deprived areas were more likely to refuse later on when visited in person by a field interviewer. If replicated in other studies, one implication would be that study designs recruiting participants only on the basis of passive/presumed consent (i.e. individuals are included in the survey unless they actively opt out) might experience an *over*-representation of individuals from deprived areas. Such designs have been common for surveys of children (with school-based surveys assuming parental consent unless parents respond to an opt-out letter) but are also widely used in adult surveys. In this context, estimating deprivation-related response biases may be important not only for making valid inferences from individual studies, but also for interpreting discrepancies between studies using active and passive consent strategies. This also highlights that the need to measure non-response by area deprivation directly from the data if one is to use weights to ‘correct’ (albeit partially and imperfect) for selective participation, as the weights cannot be assumed to be similar across study designs.

A further novel aspect of this study was that it examined non-participation rates not only for the primary adult informant (parents) but also for two other informants (teachers and children) for whom the parents acted as ‘gatekeepers’. It is striking that even when parents had agreed to take part and had agreed that the teacher/child be approached to take part, both teachers and children *still* showed higher rates of non-participation in more deprived areas. This did not seem to be mediated by child mental health problems, suggesting that this finding may be relevant to surveys using outcomes other than mental health. If so, this raises the possibility of a ‘multiplication’ of deprivation-related participation biases operating at two distinct points on the recruitment process in surveys in which some informants are only approached after obtaining consent from other informants.

## **Strengths and limitations**

The current study has several strengths, including the use of large, nationally-representative surveys; the use of an up-to-date measure of area deprivation based on a sophisticated theoretical model and extensive measurement; and the availability, at least for parents, of some detail as to why non-participating individuals did not take part. There are also several important limitations. One is the necessity of imputing IMD decile for the 7.6% of children whose parents refused in advance to the CBR. Yet while this means that our finding of higher rates of refusal in advance must be treated with some caution, two lines of evidence increase our confidence that this may be a genuine effect. First, the finding was replicated with respect to refusal rates to the second opt-out letter from ONS, for which all IMD scores are known. Second, while sensitivity modelling suggests that this result is relatively sensitive to the possibility of an ecological fallacy in which it was the more deprived individuals in more advantaged areas who refused in advance to the CBR, multi-level modelling of the other non-participation variables provides no evidence that such an ecological fallacy is likely to apply. Multilevel modelling for the child and teacher outcomes using family-level socioeconomic deprivation likewise provides no evidence that an ecological fallacy is likely to apply within the small areas over which IMD are calculated.

Our study was also limited with regard to the level of detail with which we could explore non-participation rates in teachers and children. In cases when the parent did give consent to approach teachers or children, we either know nothing (for teachers) or relatively little (for children) about why those teachers and children did not take part. With regard to these informants, therefore, our findings need to be considered as preliminary data only, which we hope will motivate future researchers to attempt to record and understand reasons for non-response in more detail. Such research is particularly important because although teachers and children show the same pattern of non-response by area deprivation as parents, the *reasons* for this selective participation may differ. For example, while the non-participation gradient for parents was primarily driven by non-contact this does not mean that the teacher gradient was likewise driven by teachers of children living in more deprived areas being less likely to receive our letters. The vast majority of participating and consenting parents (>99%) provided a valid school address to which both an initial and a follow-up reminder letter were sent. As such, it seems plausible that most teacher non-response is due to their failure to return the questionnaires rather than failure to receive the letter.

To summarise, increasing area deprivation was associated with higher non-participation across the deprivation spectrum, not only in the primary adult informant (parents) but also in children and teachers. The size and even direction of the bias differed, however, for different types of non-participation. The under-representation of parents from deprived areas was primarily driven by non-contact rather than by non-cooperation, and parents from deprived areas were in fact *over*-represented when opt-out methods were used. If replicated, these results suggest that to minimise deprivation response studies should use active consent; put effort into minimising non-contact rates; and avoid gatekeeper recruitment strategies wherever feasible/ethical. Even if some of these specific recommendations are not replicated, or are not feasible/ethical for particular studies, we suggest that all surveys using postcode-based sampling frames consider assessing empirically the degree of response bias by area deprivation. More generally, we believe our findings illustrate why it may be useful –both for refining recruitment design and for interpreting results – to record detailed information about who is not taking part in your surveys and why.

## **Competing interests**

None

## **Acknowledgements**

Many thanks to Dave Leon and Robert Goodman for their input in the design of this project, and their comments drafts of this paper. Thanks also to Mike Kenward for advice given on the statistical methods used. The B-CAMHS surveys were carried out by the Office for National Statistics, and paid for by the Department of Health, the Scottish Health Executive, and the National Assembly for Wales (1999 survey only). AG carried out the analyses for this paper while funded by an Economic and Social Research Council/Medical Research Council Research Degree Studentship.

## **References**

1. Beerten R. The effect of interviewer and area characteristics on survey response rates: an exploratory analysis. *Survey Methodology Bulletin*. 1999;45:7-15.
2. Groves M, Couper R. *Nonresponse in household interview surveys*. New York: Wiley Interscience; 1998.
3. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD. Response rates among control subjects in case-control studies. *Ann Epidemiol*. 1995 May;5(3):245-9.
4. Martin J, Matheson J. Responses to declining response rates on government surveys. *Survey Methodology Bulletin*. 1999;45:33-7.
5. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol*. 2006 Feb 1;163(3):197-203.
6. Plewis I, Ketende S. *Millennium Cohort Survey first survey: Technical report on Sampling*, 3rd edition. London: Centre for Longitudinal Studies, Institute of Education; 2004.
7. Plewis I, Ketende S. *Millennium Cohort Survey: Technical report on Response*. London: Centre for Longitudinal Studies, Institute of Education; 2006.
8. Smith AG, Fear NT, Law GR, Roman E. Representativeness of samples from general practice lists in epidemiological studies: case-control study. *Bmj*. 2004 Apr 17;328(7445):932.
9. UK Childhood Cancer Study Investigators. The United Kingdom Childhood Cancer Study: objectives, materials and methods. *Br J Cancer*. 2000 Mar;82(5):1073-102.
10. Angus VC, Entwistle VA, Emslie MJ, Walker KA, Andrew JE. The requirement for prior consent to participate on survey response rates: a population-based survey in Grampian. *BMC Health Serv Res*. 2003 Nov 18;3(1):21.
11. Tickle M, Milsom KM, Blinkhorn AS, Worthington HV. Comparing different methods to detect and correct nonresponse bias in postal questionnaire studies. *J Public Health Dent*. 2003 Spring;63(2):112-8.
12. Medical Research Council. *MRC Ethics Guide: Medical research involving children*. 2004 [cited 11th March 2008]; Available from: <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002430>
13. Leon DA, Saburova L, Tomkins S, McKee M, Shkolnikov VM. Alcohol consumption and public health in Russia. *Lancet*. 2007 Aug 18;370(9587):561.

14. Green H, McGinnity A, Meltzer H, Ford T, Goodman R. Mental health of children and young people in Great Britain, 2004: Palgrave MacMillan; 2005.
15. Meltzer H, Gatward R, Goodman R, Ford T. Mental health of children and adolescents in Great Britain. London: The Stationery Office; 2000.
16. Ford T, Goodman R, Meltzer H. The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry*. 2003 Oct;42(10):1203-11.
17. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry*. 1997 Jul;38(5):581-6.
18. The American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Lenexa, Kansas: The American Association for Public Opinion Research; 2000.
19. Office of the Deputy Prime Minister. The English Indices of deprivation 2004 - summary (revised). 2004 [cited 26 Feb 2008]; Available from: <http://www.communities.gov.uk/documents/communities/pdf/131209>
20. Office of the Deputy Prime Minister. Indices of deprivation 2004 - Population Denominators. 2007 [cited 26th Feb 2008]; Available from: <http://www.communities.gov.uk/archived/general-content/communities/indicesofdeprivation/216309/indicesdeprivation/>
21. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Statist Assoc*. 1986;81:366-74.

## **APPENDIX**

### **Supplementary information 1: Multiple imputation process**

For the children whose IMD score was missing because they had no valid postcode, we performed multiple imputation using an approximate Bayesian Bootstrap (1). The underlying model for this imputation method involves two independent stages: first a population of possible IMD deciles is created by sampling at random and with replacement the observed distribution in the postal sector; second values for the missing individuals are imputed by sampling at random and with replacement from the newly created population of IMD values. The relatively high intraclass correlation of 0.71 for IMD deciles within postcode sectors supports the imputation of IMD deciles in this way. It should be noted, however, that this model assumes that within a given postcode sector individuals were missing completely at random with respect to their IMD decile. Multiple imputation was performed using the ‘hotdeck’ command in STATA and by combining twenty-five imputations for each analysis.

### **Supplementary information 2: Sensitivity analyses regarding the validity of the imputation model**

The IMD scores of parents who refused in advance to the CBR are imputed under the assumption that within any given postal cluster the individuals who opt out are ‘missing completely at random’. We conducted sensitivity analyses assuming that those who refused in advance to the CBR were drawn systematically either from the most deprived 75% of scores in their postal sector or from the least deprived 75%. The overall effect of deprivation on parent non-response was robust to these changes: under both models, the point estimate and confidence intervals for the effect of deprivation on non-response remained unchanged to two decimal places. By contrast, the relationship between refusal in advance and area deprivation was very sensitive even to small changes: from an estimated OR of 0.96 (95% CI 0.92 – 0.98), the gradient increased to 0.91 (95% CI 0.89 – 0.92) when sampling from the least deprived 75%, and reduced to 1.01 (95% CI 0.99 – 1.102) when sampling from the most deprived 75%. For refusal in advance to the CBR - the non-participation category for which area deprivation was systematically missing - sampling from the most deprived 75% reversed the sign of the gradient from an estimated OR of 0.97 (95% CI 0.92 – 0.98) in the total sample to 1.03 (95% CI 1.01 – 1.105).

As such, the point estimate of the gradient of the relationship between refusal in advance and IMD should certainly be treated with caution. Of arguably greater concern than the accuracy of the absolute value is the possibility that the sign of the gradient could be incorrect if it were specifically more deprived individuals within less deprived postal sectors who were refusing in advance. To explore this, we used SAS to fit multilevel random effects models, levels in question were Lower Super Output Areas (the level at which IMD scores are calculated), nested within Postal Sectors (the level from which missing IMD scores were imputed). We did this for the 11 parent, teacher and child non-participation variables in Tables 1, 2 and 3 other than refusal in advance. We also fitted this model to the non-participation category most similar to refusal in advance to the CBR, namely refusal in advance to the ONS.

In 5/12 cases there was some evidence that the between-cluster gradient of the effect of IMD decile on non-participation differed from the within-cluster gradient, with the between-cluster gradient being larger in all cases. In no case, however, was the sign of the regression coefficient different between postal sectors and within postal sectors. This included refusal in advance to the ONS, for which both within- and between-cluster gradients were negative and where there was no evidence of a difference between them ( $p=0.33$ ).

### **Supplementary information 3: Sensitivity analyses in teachers and children of the effect using of small-area deprivation not individual-level indicators of socio-economic deprivation**

We conducted sensitivity analyses to explore the possibility of an ecological bias through the use of a small-area measure of deprivation rather than an individual-level measure. This was possible only for teachers and children, for whom we had parent-reported information about family socio-economic position for participants and non-participants alike. We created a continuous latent variable measuring family socioeconomic deprivation based on maternal qualification, household income, living in rented accommodation and whether either parent was economically active. The latent variable was created by fitting a multivariate probit analysis (2) using the extension to ordinal data of a one-parameter normal ogive item response model (3) and loading all four indicators of SEP onto a single factor. The Comparative Fit Index, Tucker Lewis Index and Root Mean Square Error of Approximation all indicated good model fit (with values of 0.999, 0.998 and 0.038 respectively).

The Pearson’s correlation between IMD decile and this individual-level measure of family socioeconomic deprivation was 0.51. We repeated the logistic regression analyses reported in Tables 2 and 3 but replacing IMD decile with family socioeconomic deprivation. In all 7 cases there was evidence of higher non-participation being associated with greater socioeconomic deprivation ( $p<0.001$  for 6/7 variables and  $p=0.03$  for ‘parent refused consent to approach teacher’). To explore whether there might be an ecological fallacy within the Lower Super Output Area (LSOAs) across which IMD scores are calculated, we used SAS to fit multilevel random effects models to the 7 teacher and child non-participation variables. The levels in question were individuals nested within LSOAs. In 2/7 cases there was some evidence that the between-cluster gradient of the effect of IMD decile on non-participation differed from the within-cluster gradient, with the between-cluster gradient being larger in both cases. In no case, however, was the sign of the regression coefficient different between LSOAs and within LSOAs. There is therefore no evidence of an ecological bias within LSOA of a sort which would change the substantive conclusions of this paper.

### **References for supplementary information**

1. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Statist Assoc.* 1986;81:366-74.
2. Muthen B. Latent Variable Structural Equation Modelling with Categorical-Data. *Journal of Econometrics.* 1983;22(1-2):43-65.
3. Muthen B. A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika.* 1984;49(1):115-32.