

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



Morris-Cunnington, MC; Edmunds, WJ; Miller, E; Brown, DWG  
(2004) A novel method of oral fluid collection to monitor immunity to  
common viral infections. *Epidemiology and infection*, 132 (1). pp. 35-  
42. ISSN 0950-2688 DOI: <https://doi.org/10.1017/s0950268803001493>

Downloaded from: <http://researchonline.lshtm.ac.uk/14847/>

DOI: [10.1017/s0950268803001493](https://doi.org/10.1017/s0950268803001493)

#### Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: Copyright the publishers

---

## A novel method of oral fluid collection to monitor immunity to common viral infections

---

M. C. MORRIS-CUNNINGTON<sup>1,2\*</sup>, W. J. EDMUNDS<sup>1</sup>, E. MILLER<sup>1</sup>  
AND D. W. G. BROWN<sup>3</sup>

<sup>1</sup> *Immunisation Division, Communicable Disease Surveillance Centre (CDSC), Health Protection Agency (formerly Public Health Laboratory Service)*

<sup>2</sup> *Infectious Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine*

<sup>3</sup> *Enteric, Neurological and Respiratory Laboratory, Central Public Health Laboratory, Health Protection Agency*

(Accepted 28 July 2003)

### SUMMARY

Serological surveys among representative population samples have proved rare given their reliance on invasive sample collection. We therefore completed the first population-based postal survey of immunity in England and Wales using new oral fluid technology. This paper examines the feasibility of this new methodological approach. Nearly 5500 oral fluid samples were collected, with individual demographic and social data via a questionnaire, from persons under 45 years of age recruited through general practices. Instructions were accurately followed with only 1% of samples returned without risk-factor data. The overall response rate was 40%. Response was independently associated with age, sex and location. Response was highest in children aged 5–14 years, adult females and in rural locations. This approach allowed the successful collection of comprehensive individual risk data, but response rates in adults must be improved if oral fluid surveys are to routinely complement serological surveillance.

### INTRODUCTION

Data on the proportion of the population that is immune or has been infected with a specific virus has many important epidemiological applications. These include the identification of susceptible groups in the population, the evaluation of vaccine uptake and efficacy of other health programmes and use of these data in mathematical modelling to predict outbreaks and transmission patterns [1].

Surveys of immunity among representative samples of the general population are rare due mainly to recruitment problems given the invasive nature of blood collection. This led to a reliance on anonymized,

opportunistic serum collections which have little risk-factor data attached and questionable generalizability to the national population [2].

This paper describes the first population-based survey of immunity to common viral infections using oral fluid collected by post in England and Wales. This study exploited the availability of new assays able to detect antibody markers in minimally invasive oral fluid samples and aimed to estimate the antibody prevalence of, and identify risk factors for, immunity and past infection with four common viral infections [hepatitis A virus (HAV), Epstein–Barr virus (EBV), herpes simplex virus type 1 (HSV-1) and varicella zoster virus (VZV)]. This paper concentrates on the feasibility and cost of this approach as a complement to current routine serosurveillance through a review of the study logistics and the overall response rates.

\* Author for correspondence: M. C. Morris-Cunnington, Glaxo-SmithKline, New Frontiers Park South, 3rd Avenue, Harlow, Essex, CM19 5AW.

## METHODS

### The sampling frame

The sampling frame was the Medical Research Council (MRC) General Practice Research Framework (GPRF) which allowed individuals of all ages to be sampled through practice registers. The GPRF currently consists of 1060 group general practices representing 9% of all UK practices and approximately 12% of the UK population. The composition of the GPRF does not mirror that of UK general practices overall, but there are sufficient practices of all types, and in all areas to provide representative samples where necessary [3].

A stratified, clustered study design was used. During recruitment, practices were stratified by location [North, Midlands, South, according to the Royal College of General Practitioners (RCGP) Sentinel Surveillance Scheme, and London as defined by Greater London Area health authority boundaries] [4]. Each location was then divided into Carstairs tertiles to ensure a range of area-based social deprivation scores (tertiles based on 1991 data for electoral wards). The Carstairs deprivation measure was chosen as it is more closely related to social class (through unemployment and car ownership) than other measures such as the Jarman score [5, 6]. Between 2 and 4 practices were recruited within each of the 12 strata. Fewer practices were recruited from London relative to the lower population, and lower social strata were over-sampled as those areas are often less well represented in postal population surveys [7, 8].

Practice recruitment was carried out by the MRC GPRF coordinating centre with targeted approaches to fulfil the location and deprivation criteria.

### Sample size

Sample size calculations were performed using equation (1) in Microsoft Excel (Microsoft, USA) [9, 10]. The hypothesis was that the age-specific prevalence of HAV and EBV were the same as in recent UK population-based seroprevalence surveys (1996 and 1994 respectively)  $\pm 1.5\text{--}3.0\%$  [11, 12]. Higher precision was required for the 0–4 and 15–19 year age groups which may represent peaks in viral transmission [11–13]. A design effect was incorporated into calculations to take account of the ratio of the variance of prevalence estimates assuming simple random sampling and incorporating the clustered design. This needed an estimate of between practice variation in

seroprevalence [s.d. in equation (1)] which was extrapolated from a 1996 multi-centre HAV serosurvey (age-specific design effects ranged from 1.2 to 2.0) [11].

The absolute precision for varying sample sizes within the clusters was calculated as follows:

$$\begin{aligned} \text{absolute precision } (\pm \%) \\ = 2 \times \sqrt{(\text{s.e.}^2 + \text{s.d.}^2) / \text{number of clusters}}, \end{aligned} \quad (1)$$

The standard error (s.e.) is within

$$\begin{aligned} \text{practice variation} \\ = \sqrt{[\text{prevalence} \times (1 - \text{prevalence})] / N}, \end{aligned} \quad (2)$$

with  $N$  as the practice sample size. The standard deviation (s.d.) is between practice variation for a prevalence estimate.

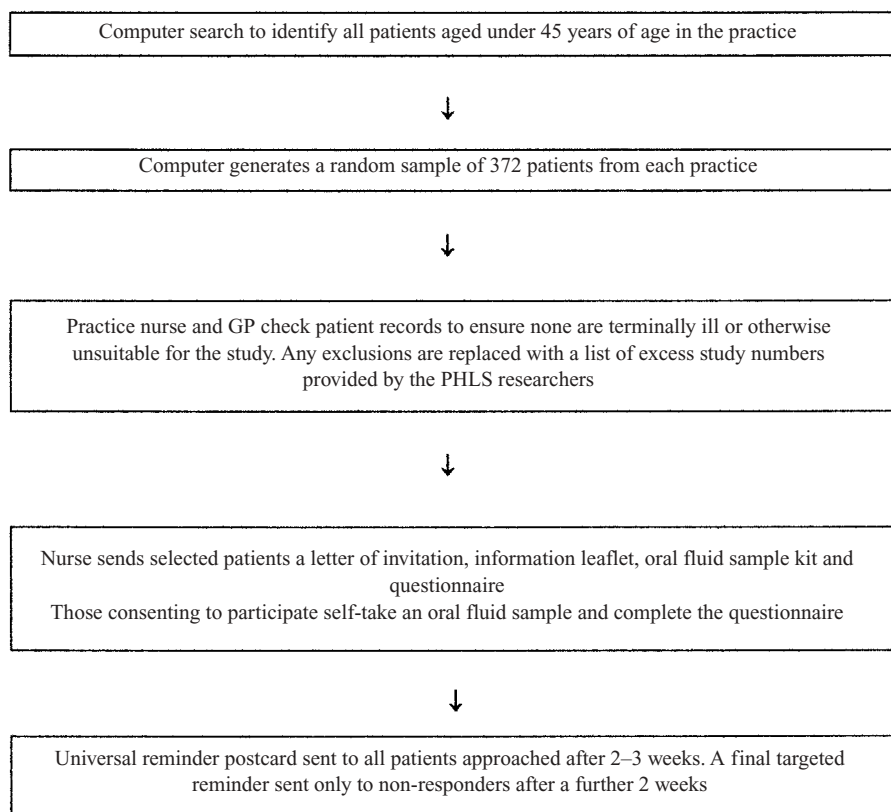
The calculations gave a sample size of 168 individuals per practice. Assuming a 50% response rate and that 10% of register addresses were inaccurate (due to death or movement), figures were adjusted to 372 per practice [14–16]. A cluster number of 40 was chosen as the figure that minimized the overall sample size while remaining a realistic recruitment goal. This gave a total sample size of 14 800 individuals.

### Patient selection

Patient selection, sample and data collection took place between September 2001 and June 2002. Recruitment was limited to individuals aged under 45 years, as after that age there are few new infections and little change in the relevant seroprevalence profiles [11, 12]. The randomized patient selection and recruitment procedures, finalized after a pilot study [15], are described in Figure 1.

To encourage participation the study was advertised through posters in the practice waiting rooms. The letters of invitation were printed on practice headed notepaper and were signed by the GP. The information sheets also contained the official logos of the Public Health Laboratory Service (PHLS) and the London School of Hygiene and Tropical Medicine (LSHTM) as one meta-analysis of postal surveys identified university and government sponsorship as a factor positively affecting response rate [17].

If the practice nurse or GP believed that a substantial proportion of the practice sample were non-English speakers, all correspondence was translated into the relevant language. This was only necessary



**Fig. 1.** Flow chart of study design.

for one practice in north London where it was estimated that 25% of the register were Turkish and spoke very poor English.

Different letters, information sheets and questionnaires were provided for adults and for those aged under 16 years. Correspondence for the latter was addressed to the child's guardian. Children aged 8–15 years were provided with an additional information sheet aimed specifically at their age group so that they could be involved in the decision of whether or not to participate in the study.

The selected individuals were asked to collect an oral fluid sample using the Oracol sponge device (Malvern Medical Developments, Worcestershire, UK) and to fill in a short questionnaire collecting demographic and social data. The questionnaires were shortened versions of the pilot version [15]. The questions were mostly multiple choice and were based on the 1991 census and the General Household Survey [18, 19].

All postage was prepaid. The sample kits were despatched second class, but their return was prepaid first class as it was important for the oral fluid samples to reach the laboratory as quickly as possible to minimize the possibility of antibody decay.

### **Ethical approval**

Ethical approval was granted by the North Thames Multi-Regional Ethics Committee and by the ethical committees of the PHLS and LSHTM.

### **Statistical analysis**

Data on both responders and non-responders were restricted to age group and sex as well as the Carstairs tertile, location, urban/rural location and size of the practice with which the individual was registered. These data were used to conduct an analysis of factors related to response in STATA [20].

The stratified, clustered study design was adjusted for by allocating each individual to a specific stratum, primary sampling unit (cluster/practice) and population weighting (probability of being sampled). The weighting was calculated by estimating the stratum population from census and small area statistics data (e.g. the actual population aged 0–4 years living in the north and lowest Carstairs tertile) and dividing it by the number of individuals selected within each stratum. These considerations allowed sampling estimates and their variance to be adjusted for the study design [21].

A single variable logistic regression was performed with response ('yes/no' for return of oral fluid sample) as the dependent variable to ascertain which of the explanatory variables were significant at the 10% level. Significant variables were entered into a multi-variable analysis and the significance of each one tested again at the 5% level using a likelihood ratio test. Any two-way interactions within the final model were investigated. The baseline category for each variable was taken as the lowest group (e.g. lowest age group) or the most common category.

### Cost of oral fluid sample collection

The cost per response (sample collected) was estimated for the oral fluid survey considering only consumable, postal and staff costs. These costs are outlined in Table 1. Nurse time varied greatly per practice according to the level of practice computerization (i.e. whether mail merges were possible). The average cost per practice was therefore estimated. It was assumed that all selected individuals were sent a first reminder letter and all non-responders were sent a second.

## RESULTS

### Practice recruitment

A total of 40 practices were recruited from across England and Wales. There was good agreement between the recruited and desired sample, although the north was slightly over-represented and the most deprived areas of London were under-represented (Table 2).

### Study sample

A total of 5457 oral fluid samples were returned out of the 14398 kits despatched. Fifty-four of these (0.99%) were returned without questionnaires and a further 13 were excluded for being from individuals aged over 44 years. This left a total of 5390 samples, 2452 from males and 2938 from females.

Instructions appeared to be well understood with only 54 (0.99%) of the 5457 collected samples returned without questionnaires. Question completion rates were also high with only 1.6% of questionnaires incomplete for ethnicity and 1.4% for occupation. Data on such variables had not previously been available through routine serological surveillance in England and Wales.

Table 1. *Costs per oral fluid sample*

	Cost (£)
Consumables	
Swab	0.76
Kit container and packaging	1.94
Postage	
Postage for letters, oral fluid kit and questionnaire	0.33
Postage for return of oral fluid sample and questionnaire	0.41
Reminder letters	0.19
Staff costs (per practice)	
Nurse time	550.00
GP time	25.00

Table 2. *Characteristics of recruited general practices*

Location	Carstairs tertile		
	High	Mid	Low
North	4 (3)	5 (4)	4 (4)
Midlands	3 (3)	4 (4)	4 (4)
South	3 (3)	3 (4)	4 (4)
London	3 (2)	2 (2)	1 (2)

Total of 39 clinics, one extra was recruited wherever possible.

The main demographic and social characteristics of the study sample are illustrated in Table 3 and are compared to a general population adult sample from the 2000 General Household Survey (only covers adults). The comparison indicates an under-representation of ethnic minority groups and lower social groups described through routine occupations.

### Response rates

The overall weighted response rate for specimen return was 40.0%. This was a slight increase from the crude response estimate of 37.5% and resulted from the weighting, which adjusted for the over-sampling of certain groups. As young adults with poor response rates were over-sampled, the adjustment resulted in a slight increase in estimated response. The weighting made very little difference to the age-specific response estimates, therefore only the weighted age- and sex-specific response estimates are shown in Figure 2. The overall survey design effect was 3.0, slightly higher than the proxy estimate used for the sample size calculations, which could have led to some loss of precision for the prevalence estimates.

Table 3. Social and demographic characteristics of the survey population

Characteristic	Number (all)	Percentage (all)	Percentage in adults (16–44 years)	Percentage according to 2000–2001 General Household Survey (GHS) (16–44 years old only)
<b>Household size</b>				
1–2	749	13.9	22.8	No data for desired age group
3–4	3376	62.5	58.6	
5+	1180	21.9	16.8	
Unknown	86	1.7	1.8	
<b>Occupational class</b>				
Managerial and professional	2157	40.0	27.7	26
Intermediate	387	7.2	10.0	27
Routine/manual and small employers	1705	31.6	30.6	47
Students	800	14.8	26.2	No information
Not working/unknown	342	6.3	5.5	
<b>Accommodation</b>				
Owner occupied	4149	77.0	74.0	62
Rented from council	584	10.8	10.2	21
Rented from private landlord	380	7.1	8.7	16
Other or unknown	278	5.2	7.1	0
<b>Ethnicity</b>				
White	5071	94.1	94.2	92
Black (African, Caribbean, other)	33	0.6	0.80	2
South Asian (Indian, Pakistani, Bangladeshi)	28	0.5	0.40	3
Mixed race/other/unknown	259	4.8	4.5	2

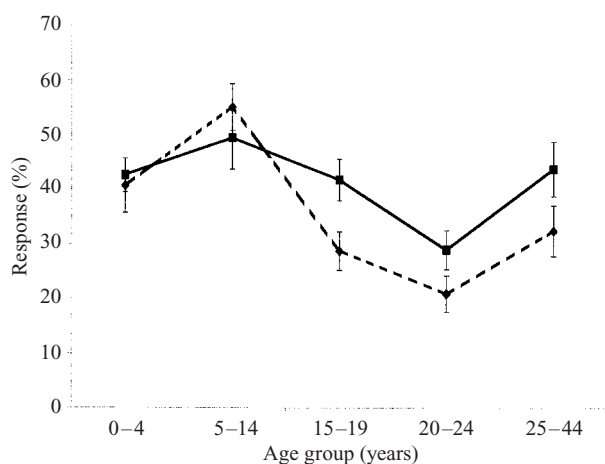


Fig. 2. Age and sex-specific weighted response rates (with 95% confidence interval). - -◆- -, Male; —■—, female.

The response rate was significantly higher in the under 15 years age group than in those aged 15 years and over (48.9 vs. 35.8%,  $\chi^2=224$ ,  $P<0.001$ ). The response rate was similar in males and females below 15 years of age ( $\chi^2=6.02$ ,  $P=0.16$ ), but in adults was consistently higher in females ( $\chi^2=117.5$ ,  $P<0.001$ ). The response among young male adults aged 20–24 years was particularly poor (~20%). Table 4

illustrates other notable trends in response, with the weighted response rate in London over 4% lower than the next lowest score (for the Midlands). There was a significant difference in response in urban and rural locations with the weighted response rate (for all ages) 10% higher in the latter ( $P<0.001$ ).

The multivariable regression model indicated that location was no longer significant at the 5% statistical level ( $F=2.23$ ,  $P=0.09$ ) and it is likely that the urban nature of London explained the initial location effect. There was also a significant sex–age group interaction ( $F=9.04$ ,  $P<0.001$ ) (see Fig. 2 and Table 5). The age-specific variation in response rates was much less pronounced in females, with only the 5–14 and 20–24 years age groups exhibiting response rates significantly different from the 0–4 years group [OR=1.32 (1.01–1.71) and OR=0.55 (0.45–0.67) respectively]. In males all adult age groups had response rates significantly lower than the youngest age group, apart from the 5–14 years group, where response rates were significantly higher [OR=1.78 (1.39–2.28)] (Table 5). The age- and sex-adjusted odds of response in individuals in both mixed and urban areas were approximately 35% less than those from rural areas [mixed vs. rural OR=0.64 (0.51–0.80) and urban vs. rural OR=0.66 (0.55–0.78)].

Table 4. Crude and weighted response rates according to demographic data

	Crude % response	Weighted % response	$\chi^2$ value for overall significance of variable
<b>Sex</b>			
Male	34.6	36.8	$\chi^2 = 31.2$ , $P < 0.001$
Female	40.3	43.2	
<b>Age group (years)</b>			
0-4	42.9	41.6	$\chi^2 = 39.3$ , $P < 0.001$
5-14	51.0	52.2	
15-19	35.8	35.3	
20-24	25.3	24.9	
25-44	38.6	38.1	
<b>Carstairs tertile</b>			
Low	38.8	42.7	$\chi^2 = 1.7$ , $P = 0.2$
Mid	37.2	40.2	
High	36.6	38.4	
<b>Location</b>			
North	40.9	42.5	$\chi^2 = 4.6$ , $P = 0.004$
Midlands	36.6	38.9	
South	36.5	40.8	
London	33.7	34.6	
<b>Urban/rural</b>			
Rural	41.2	47.4	$\chi^2 = 7.8$ , $P = 0.001$
Mixed	35.6	37.2	
Urban	36.7	37.8	
<b>Practice size</b>			
< 5000	38.0	41.0	$\chi^2 = 1.4$ , $P = 0.3$
5000-9999	38.7	42.8	
10 000-14 999	35.9	37.8	
15 000+	36.9	36.6	

### Cost of oral fluid sample collection

A total of 14 398 individuals were sent oral fluid kits at a cost of £3.03 per person. All individuals were sent a first reminder letter (£0.19 × 14 398) and 60% of those a second (£0.19 × 8639). A total of 5457 oral fluid samples were returned at a cost of £0.41. Nurse and GP time in the 40 practices accounted for £23 000. This gave an overall cost of £71 003 with 5457 responses and a cost per sample of £13.01.

## DISCUSSION

To our knowledge this was only the second population-based survey of immunity to viral infections using a postal oral fluid collection anywhere, and the first to collect extensive individual demographic and

Table 5. Sex specific odds ratios (95% CI) from a multivariable logistic regression analysis of response rates according to age and urban/rural location

Variable	Odds ratio (95% CI) for males	Odds ratio (95% CI) for females
<b>Age group (years)</b>		
0-4	1.0	1.0
5-14	<b>1.78 (1.39-2.28)</b>	<b>1.32 (1.01-1.71)</b>
15-19	<b>0.58 (0.45-0.76)</b>	0.96 (0.78-1.19)
20-24	<b>0.38 (0.29-0.50)</b>	<b>0.55 (0.45-0.67)</b>
25-44	<b>0.69 (0.53-0.88)</b>	1.04 (0.83-1.29)
<b>Sex by age group</b>		
0-4	1.0	1.08 (0.89-1.31)
5-14	1.0	0.79 (0.62-1.02)
15-19	1.0	<b>1.78 (1.44-2.20)</b>
20-24	1.0	<b>1.56 (1.15-2.11)</b>
25-44	1.0	<b>1.63 (1.41-1.89)</b>

Figure 2 shows these results as response rates.

Figures in bold refer to odds ratios significantly different from the baseline.

social data in relation to antibody prevalence data. The survey approach proved feasible with the collection of approximately 5400 samples. Instructions were followed well, ensuring that the prevalence of antibody markers could be interpreted in relation to a wide range of demographic and social data which were not previously available through routine serological surveillance in England and Wales. However, the representativeness of the generated sample remains questionable due to potential sources of selection bias, which must be considered.

The overall response proved disappointing at 40%. Non-response, associated with young adult age, male gender and urban location, was probably a mix of true refusals and practice register address inaccuracies, which have been estimated at between 3 and 15% of the register in previous studies [16, 22, 23]. Differential response by age and sex is of particular concern given the variation in the prevalence of viral antibodies due to age- and sex-specific differences in the chance of exposure or vaccination, and time since exposure. Such biases in seroprevalence estimates can be partially corrected through statistical weighting for differential response, but further corrections are only possible if data are available on seroprevalence in non-responders. This is rare.

The low response among young adults (~20%) was expected from the pilot and other studies [14, 15]. Young adults are a highly mobile group: the annual

British Household Panel Study reported that in 2000–2001 approximately 50% of people aged 16–24 years had been at their address for less than a year illustrating high rates of movement [24]. However, a failure to aggressively target recruitment among young adults was a limitation of this study. An alternative approach was decided upon to provide supporting data. This included additional, targeted surveys of young adult groups (e.g. in universities, youth clubs, prisons). A subsequent opportunistic cross-sectional study at Coventry University proved successful with the collection of nearly 1000 oral fluid samples from individuals aged 18–25 years within a few days [25]. However, there is a need to investigate other methods, such as incentives, to increase young adult recruitment within the main survey design.

The target response rate of 50% was reached in children aged 5–14 years making this approach practicable in children. This may reflect an increased likelihood of registration with a GP at a young age (with greater address accuracy).

The sex response differential was only present in adults and may reflect different approaches to health matters with females placing a higher priority on such issues increasing their willingness to participate in health research [26, 27]. The similar sex response rates in children probably reflected the involvement of parents in the participation decision, treating children of both sexes equally.

The survey response rates proved disappointing compared to the figure of 60% achieved in the only other population-based postal survey of viral immunity (to hepatitis B virus) using oral fluid samples [28, 29]. This Irish study used a four-letter approach (initial warning letter, invitation letter and two follow-up letters) to recruit households as well as a final telephone reminder and a prize incentive. The higher response rate could be due to the intensive follow up or the incentive, although these come at an extra cost which was one of the main limiting factors in our study. Telephone contact is also increasingly difficult in the United Kingdom due to data protection restrictions.

Comparisons between the studies are complex as the Irish Study did not collect demographic data beyond age and sex, and employed a different sampling frame: the electoral roll. Further work is now required to ascertain the greatest response influencing factor: the sampling frame, the initial warning letter, intensive follow up, telephone follow up, prize incentives or length of questionnaire. Pilot studies could

establish the importance of each factor (e.g. response in individuals receiving an incentive or not) and could contribute to better study design.

The first postal population-based survey of immunity to common infections using oral fluid samples in England and Wales was feasible and allowed the collection of comprehensive risk-factor data. The sampling approach was most successful in children and the collection of additional demographic and social data, not previously available, was achieved. However, the response rates in adults were well below the target of 50% and differential response put the representativeness of the sample into doubt. The use of other population sampling frames, or more targeted approaches through universities, employment or youth centres should be investigated further to ascertain whether postal collections in adults could prove a cost-effective and scientifically valid complement to routine serological surveillance.

#### ACKNOWLEDGEMENTS

The authors are grateful to Madge Vickers and Jeanett Martin for allowing access to the MRC GPRF and for comments on the initial study proposal. This study would not have been possible without the tireless recruitment efforts of Nicky Fasey of the MRC and all the practice nurses who gave up their time to address and send out thousands of test kits and questionnaires. Finally the authors thank Nick Andrews and Phillippa Cumberland for valuable statistical support and advice. This study was funded by the Public Health Laboratory Service.

#### Conflict of interest

Marianne Morris-Cunnington now works for Glaxo-SmithKline, but completed this survey and initial drafts of this paper while still working at CDSC and the London School of Hygiene and Tropical Medicine.

#### REFERENCES

1. Gay NJ, Hesketh L, Morgan-Capner P, Miller E. Interpretation of serological surveillance data for measles using mathematical models: implications for vaccine strategy. *Epidemiol Infect* 1995; **115**: 139–156.
2. Osborne K, Gay NJ, Hesketh L, Morgan-Capner P, Miller E. Ten years of serological surveillance in England and Wales: methods, results, implications and action. *Int J Epidemiol* 2000; **29**: 362–368.
3. The Medical Research Council General Practice Research Framework. 2003 ([www.mrc-gprf.ac.uk](http://www.mrc-gprf.ac.uk)).



4. Fleming DM. Weekly returns service of the Royal College of General Practitioners. *Comm Dis Pub Health* 1999; **2**: 96–100.
5. Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland and England and Wales. *BMJ* 1989; **299**: 886–889.
6. Carstairs V. Deprivation indices: their interpretation and use in relation to health. *J Epidemiol Comm Health* 1995; **49**: S3–S8.
7. Melton 3rd LJ, Dyck PJ, Karnes JL, O'Brien PC. Non response bias in a study of diabetic complications: The Rochester Diabetic Neuropathy Study. *J Clin Epidemiol* 1993; **46**: 341–348.
8. Risk in communities (ARIC) study investigators. The effect of non response on prevalence estimates for a referent population: insights from a population-based cohort study. *Ann Epidemiol* 1996; **6**: 498–506.
9. Lemeshow S, Hosmer DW, Klar J, Lwanga K. Adequacy of sample size in health studies. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley and Sons for WHO, 1990.
10. Microsoft Corporation. Microsoft Excel (1997). 1997. Seattle, USA.
11. Morris MC, Gay NJ, Hesketh L, Morgan-Capner P, Miller E. The changing epidemiological pattern of hepatitis A in England and Wales. *Epidemiol Infect* 2002; **128**: 457–463.
12. Morris MC, Edmunds WJ, Hesketh L, et al. Sero-epidemiological patterns of Epstein–Barr and herpes simplex (HSV-1 and HSV-2) viruses in England and Wales. *J Med Virol* 2002; **67**: 522–527.
13. Gay NJ, Morgan-Capner P, Wright J, Farrington P, Miller E. Age-specific antibody prevalence to hepatitis A in England: implications for disease control. *Epidemiol Infect* 1994; **113**: 113–120.
14. De Melker HE, Nagelkerde NJD, Conyn-Van Spaendonck MAE. Non-participation in a population-based seroprevalence study of vaccine-preventable diseases. *Epidemiol Infect* 2000; **124**: 255–262.
15. Morris MC, Edmunds WJ, Miller E, Brown DWG. Oral fluid collection by post – a pilot study of two approaches. *Public Health* 2002; **116**: 113–119.
16. Sethi D, Wheeler JG, Cowden JM, et al. A study of infectious intestinal disease in England: plan and methods of data collection. *Comm Dis Public Health* 1999; **2**: 101–107.
17. Fox JR, Crask MR, Kim J. Mail survey response rate: a meta-analysis of selected techniques for including response. *Public Opin Q* 1988; **52**: 467–491.
18. The General Household Survey. 2003 ([www.statistics.gov.uk/lib](http://www.statistics.gov.uk/lib)).
19. Office of population censuses and surveys. 1991 census. Definitions Great Britain. London: HMSO, 1992.
20. Stata Corporation. Stata Statistical Software (6.0). 1999. College Station, Texas, USA.
21. Eltinge JL, Scribney WM. Some basic concepts for design-based analysis of complex survey data. *Stata Tech Bull* 1996; **31**: 3–6.
22. O'Mahony PG, Thomson RG, Rodgers H, Dobson R, James OFW. Accuracy of the family health services authority register in Newcastle upon Tyne, UK as a sampling frame for population studies. *J Epidemiol Comm Health* 1997; **51**: 206–207.
23. Roe L, Mant D, Coulter A. Screening: the inadequacy of population registers. *BMJ* 1989; **298**: 1100.
24. British Household Panel Survey. 2003 ([www.statistics.gov/statbase/ssdataset](http://www.statistics.gov/statbase/ssdataset)).
25. Morris MC. The application of novel oral fluid technology to age-specific surveys of immunity. 2003. London School of Hygiene and Tropical Medicine.
26. Mishra SI, Dooley D, Catalano R, Serxner S. Telephone health surveys: potential bias from non-completion. *Am J Public Health* 1993; **83**: 94–99.
27. Cottler LB, Zipp JF, Robins LN, Spitznagel EL. Difficult to recruit respondents and their effect on prevalence estimates in an epidemiologic survey. *Am J Epidemiol* 1987; **125**: 329–339.
28. O'Connell T, Thornton L, O'Flanagan D, et al. Oral fluid collection by post for viral antibody testing. *Int J Epidemiol* 2001; **30**: 298–301.
29. O'Connell T, Thornton L, O'Flanagan D, et al. Prevalence of hepatitis B anti-core antibody in the Republic of Ireland. *Int J Epidemiol* 2000; **125**: 701–704.