

Refereed paper

Hospital data may be more accurate than census data in estimating the ethnic composition of general practice populations

Sally A Hull FRCGP MRCP MSc
Senior Lecturer

Carol Rivas BSc MSc
Research Fellow

Centre for Health Sciences, Barts and the London Queen Mary's School of Medicine, London, UK

Jacqui Bobby MA
Technical Consultant, Tribal Consulting, Hampshire, UK

Kambiz Boomla FRCGP MRCP MSc
Senior Lecturer

John Robson FRCGP MD
Senior Lecturer

Centre for Health Sciences, Barts and the London Queen Mary's School of Medicine, London, UK

ABSTRACT

Background Equity of service provision by age, ethnicity and sex is a key aim of Government policy in the UK. The prevalence, natural history and management of common chronic conditions, such as diabetes and hypertension, vary between ethnic groups. Developing and monitoring responsive local services requires accurate measures of ethnicity and language needs. Hence establishing the ethnic composition of GP populations is important.

Objective To compare three methods of estimating the ethnic composition of GP registered populations in three east London primary care trusts (PCTs).

Design Self-reported ethnicity, routinely collected at practice level (and considered the 'gold standard'), was compared with two indirect methods of attributing ethnicity. The indirect method currently used in the UK assigns ethnicity to GP populations based on geographical postcode attribution from the national census. A proposed alternative indirect method uses the ethnic breakdown of hospital admission data from practice lists to attribute ethnicity to the whole practice population. Comparisons were made between practice self-report recording and these two indirect methods. Bland–Altman plots

were used to assess the agreement between methods of measurement.

Results Data from 103 practices, covering 70% of the GP registered population, was used.

The hospital admission method showed better agreement with practice self-report data than the census attributed method. For white populations Bland–Altman plots showed a mean difference of 1.4% (95% CI –14.9 to 17.7) between hospital admission and practice data, and a mean difference of 12.5% (95% CI –6.2 to 31.1) between census attributed and practice data. Differences were also found for south Asian and black populations.

Conclusion Practice ethnicity measured using hospital attendance data is in closer agreement with practice recording of self-reported ethnicity than the census attribution method. Census attribution may provide misleading information on the ethnic composition of practice populations.

We recommend that healthcare commissioners change to this method of measurement when practice self-report data is not available.

Keywords: ethnicity, ethnic minority health, general practice

Introduction

Establishing equitable service provision by age, ethnicity and gender is an important aim of UK Government health policy.¹ Reductions in the disparity in health service utilisation and outcomes by different ethnic groups can be used as a key service quality marker in any health system.² The UK is one of the few European countries which officially recognises the need for ethnicity data to support service monitoring purposes, in contrast to France and Germany, for example, where restrictions on collecting such data exist.³ Notwithstanding the difficulties of recording ethnicity, and the potential for misuse of ethnicity as a determinant of health,^{4,5} developing reliable methods of recording ethnicity and language at primary care practice level is an essential first stage in the identification of disparities. This information can then be used in the development of local health policy, for the provision of responsive local services, and for assessing the provision of services to ethnically diverse populations by provider organisations.

Some major chronic diseases, such as diabetes, hypertension and coronary heart disease (CHD) are examples of conditions which have significantly different prevalence or management by ethnic group. Diabetes and CHD have much higher rates among south Asian groups, and there are higher rates of hypertension among black African and Caribbean people.^{6,7} In the UK setting the commissioners of primary health care, the PCTs, have an important role in monitoring practice performance and working with practices to improve quality. This may involve estimating the predicted prevalence by practice of chronic diseases, by reference to the demographic characteristics of general practice populations.

Developing robust measures of disease prevalence by practice requires an accurate estimate of the ethnic population at practice level. This is particularly important in urban areas, which are most ethnically diverse, and where population mobility is greatest. To date there have been low levels of self-reported ethnicity recording at general practice level in the UK^{3,8} and in other settings, such as the USA, where managed care plans often do not collect routine data on ethnicity.² In the absence of this 'gold standard' other methods have been used to estimate the ethnicity of practice populations. The most common of these is attribution of population characteristics based on data from the census. Hence if a practice has a registered population of 500 from a Super Output Area (SOA, a census associated geographic area with 1000 to 1500 residents) and the census records 30% of the population in the SOA as of white ethnicity, then 30% of the 500 will be recorded as white.⁹

The aim of the present study is to compare the current census related attribution method of estimating the ethnic composition of general practice registered populations with a method derived from hospital admission data, and to compare both these indirect methods with the gold standard of self-reported ethnicity captured at practice level.

Methods

Sources of data

The study was set in the three east London PCTs of Newham, Tower Hamlets and City and Hackney, with a combined GP registered population of 834 500 in mid 2006. In the 2001 UK census 51.3% of the population in these three PCTs was recorded as of non-white ethnic origin.¹⁰ In City and Hackney, 25.4% of the population was described as black African or black Caribbean. In Newham 21.0% were Indian or Bangladeshi and in Tower Hamlets, 33.3% were Bangladeshi. These PCTs represent some of the eight most deprived localities in Britain.¹¹ All three PCTs have supported incentives to promote the routine recording of ethnicity and language at practice level.

Practice data

The Clinical Effectiveness Group (CEG) has since 1997 collected routinely recorded, computerised GP data for annual audits on chronic disease management.^{12,13} Ethnicity is self-reported by patients at the practices, and recorded by five-byte Read code at registration, or during consultation, using the 16 categories of the 2001 UK census. Most practices used the 9i hierarchy (the 2001 census related Read code set), but where necessary we mapped the 9S hierarchy to the 9i (see Appendix, Figure 1).

Data are extracted from practice computers using the Morbidity Information Query and Export Syntax (MIQUEST).¹⁴ The latest available data, which we use here, are for the 15-month period ending 31 March 2007. The data on ethnicity, being self-reported and obtained from practice registers, may be considered as observed data in the present study. We therefore consider them the best standard we have against which to assess the other two data sources, which are both indirect measures of expected ethnicities.

Hospital Episode Statistics (HES) data

The hospital admissions data is derived from the HES database, published by the Information Centre of the

UK National Health Service and the Office for National Statistics (ONS) and compiled from millions of Finished Consultant Episodes (FCEs) of care.¹⁵ These link each hospital episode to a practice code, regardless of which hospital the patient attends. Recording of ethnicity is not complete, but is improving; for 2005/06 around 80% of FCEs had a valid ethnicity recorded. We used data for episodes completed between 1 April 2005 and 31 March 2006, the latest year for which data were available at the time of analysis. Only episodes of admission to hospital were used in the analysis. One patient could have more than one FCE in the same year but we counted these patients once only. In total, we extracted ethnicity records for around 111 000 different patients registered with practices in City and Hackney, Newham and Tower Hamlets. We then applied the proportions defined at admission by ethnic group and age band to the practice population with which they were registered. Thus if 50% of people in the HES data from a particular practice in a ten-year age band were South Asian then we assumed that 50% of the practice population in that age band were South Asian.

Census attributed data

The East London Common Information System (ELCIS) data used by PCTs are derived from the 2001 census, published by the ONS. To obtain estimates of practice ethnicity the data is proportionally allocated to practices by weighting SOA 2001 census data using the postcode distribution of practice lists. In other words, if a practice has 100 people from an SOA on their list (determined by postcode) then those 100 people have ethnicities attributed to them in the proportions derived from the census. This approach assumes that practice and SOA populations are ethnically similar. The ONS population data are adjusted at intervals between the census dates (using registrations of births and deaths, and estimates of domestic and international migration). We used data attributed to practice registers in mid-2006.

Ethnic groupings

For self-reported ethnicity, people were asked to choose one option from a restricted list of choices based on the 2001 census definitions. For this study we compared the different data sources by reference to five aggregated ethnic groupings. These were: white (British, Irish, other white); black (black African, black Caribbean, black British or other black people); South Asian (Bangladeshi, Pakistani, Indian, Sri Lankan, British Asian or other South Asian); mixed (parentage from two different ethnic groups) and other (including

Chinese, Vietnamese and other South East Asian). Percentages of total recorded ethnicity were also noted for each practice, as the sum of these five groups. Ethnicities not recorded included the categories of 'miscoded' and 'unknown'.

Statistical methods

All statistical analysis was conducted using Excel. One-hundred and three of the 156 practices in the three study PCTs, with a population of 583 586, were included in the analysis. Practices were excluded if their overall ethnicity recording rate was less than 10%. Practices were also excluded if there was no available HES or ELCIS data.

Scatter plots, with a line of equality, were used to make initial comparisons between the sources of data. As correlation measures the strength of a relation between two variables, rather than the agreement between them, we then compared the values from each source using Bland–Altman plots (difference against mean), a statistical method used to assess the agreement between two methods of clinical measurement, but also applicable to other forms of measurement.¹⁶

The main outcome is reported as a percentage difference in recorded ethnicity for each of the three aggregated ethnic groups with 95% confidence intervals.

Results

Data were accessed from 103 (66%) practices in three east London PCTs covering 583 586 (70%) of a total of 834 500 patients registered at GP practices in the three PCTs during 2006. Ethnicity was recorded in 65% of the population used for analysis. Breakdown by PCT is shown in Table 1.

Age profiles for recorded ethnicities were plotted separately for each ethnic group and compared to the age profiles of total practice lists (see Appendix, Figure 2 online at http://www.radcliffe-oxford.com/journals/J12_Informatics_in_Primary_Care/Supplementary%20Papers.htm). There was a lower proportion of white ethnicity recording in the 0–19 age band ($\chi^2 < 0.001$), but no significant difference for the other age bands.

The percentage total recorded ethnicity per practice for all ethnic groups was plotted against recorded ethnicities for each aggregate group for each practice (see Appendix, Figure 3 online at http://www.radcliffe-oxford.com/journals/J12_Informatics_in_Primary_Care/Supplementary%20Papers.htm). No pattern was seen for any of the three data sources, suggesting that there was no association between total ethnicity recording and the size of different ethnic groups recorded in the practice population.

Table 1 PCT ethnicity recording rates in the 103 practices included in the analysis (2006)

PCT	City and Hackney	Newham	Tower Hamlets
No. of practices (%)	30 (60%)	46 (70%)	27 (71%)
No. of population (%)*	195 293 (90%)	228 038 (92%)	160 255 (75%)
<i>Ethnicity recorded among practices used in the analysis</i>			
White	26%	17%	32%
Black	14%	15%	6%
South Asian	5%	24%	30%
Mixed	2%	1%	0.5%
Other	10%	4%	4%
None recorded	43%	39%	28%
Total ethnicity recorded	57%	61%	72%

* ONS Mid-year population estimates for 2006

A third check for bias in the data included a plot of practice list size against recording rates for each of the three aggregated ethnic groups. This showed no association (see Appendix, Figure 4 online at http://www.radcliffe-oxford.com/journals/J12_Informatics_in_Primary_Care/Supplementary%20Papers.htm). These analyses suggest that ethnicity recording rates are independent of characteristics such as practice size and ethnic breakdown, hence these factors were not considered further in the analysis.

The plot of census attributed versus practice derived values, with a line of equivalence, for white ethnicity (Figure 1a) shows a systematic difference between these two data sets, illustrating that for a large group of practices the white population is overestimated by the census attributed data. This suggests that this methodology is not an 'equivalent' method of measuring practice ethnicity compared to practice data. Similar systematic differences are noted for black and south Asian ethnicities (Figures 1b and 1c).

The plot of HES versus practice-derived values for white ethnicity (Figure 1a) shows better agreement between these two data sets, suggesting that these may be more equivalent measures.

The agreement between the methods of measurement was further explored using Bland–Altman plots (see Figure 2a). The plot of HES vs practice data for proportion of white ethnicity demonstrates good agreement between the two datasets, and shows no systematic variation related to the size of measurement. The mean difference (95% CI) between methods for white ethnicity is 1.39% (–14.93 to 17.71). The Bland–Altman plot for census attributed vs practice data shows systematic difference relating to the size of

measurement (supporting our previous analyses in demonstrating non-equivalence). The mean difference is 12.49% (95% CI –6.15 to 31.14).

The Bland–Altman plots (census attributed vs practice data) for Black and South Asian ethnicity show systematic variation at the extremes of measurement (Figure 2b, 2c) again suggesting non-equivalence. The mean differences in measurement for each ethnic group are summarised in Table 2.

Discussion

Study findings

Our results demonstrate that there is better agreement between HES and practice self-report data than between the census attributed and practice data. This suggests that the HES and practice self-report data may be used interchangeably with a high degree of confidence, whereas the census attributed ethnicity data are not accurate at the level of practices. The census data show systematic bias in attributing proportions of the major ethnic groups to practice populations. We also demonstrate that the HES and practice data may be used to indicate ethnicities of the practice population regardless of the absolute levels of ethnicity recording within each practice. We consider that the most plausible reason for the discrepancy between census attributed and practice self-report ethnicity data is best described by the 'ecological fallacy'.^{17,18} This describes an important source of bias in epidemiological studies in that subsets of a population will

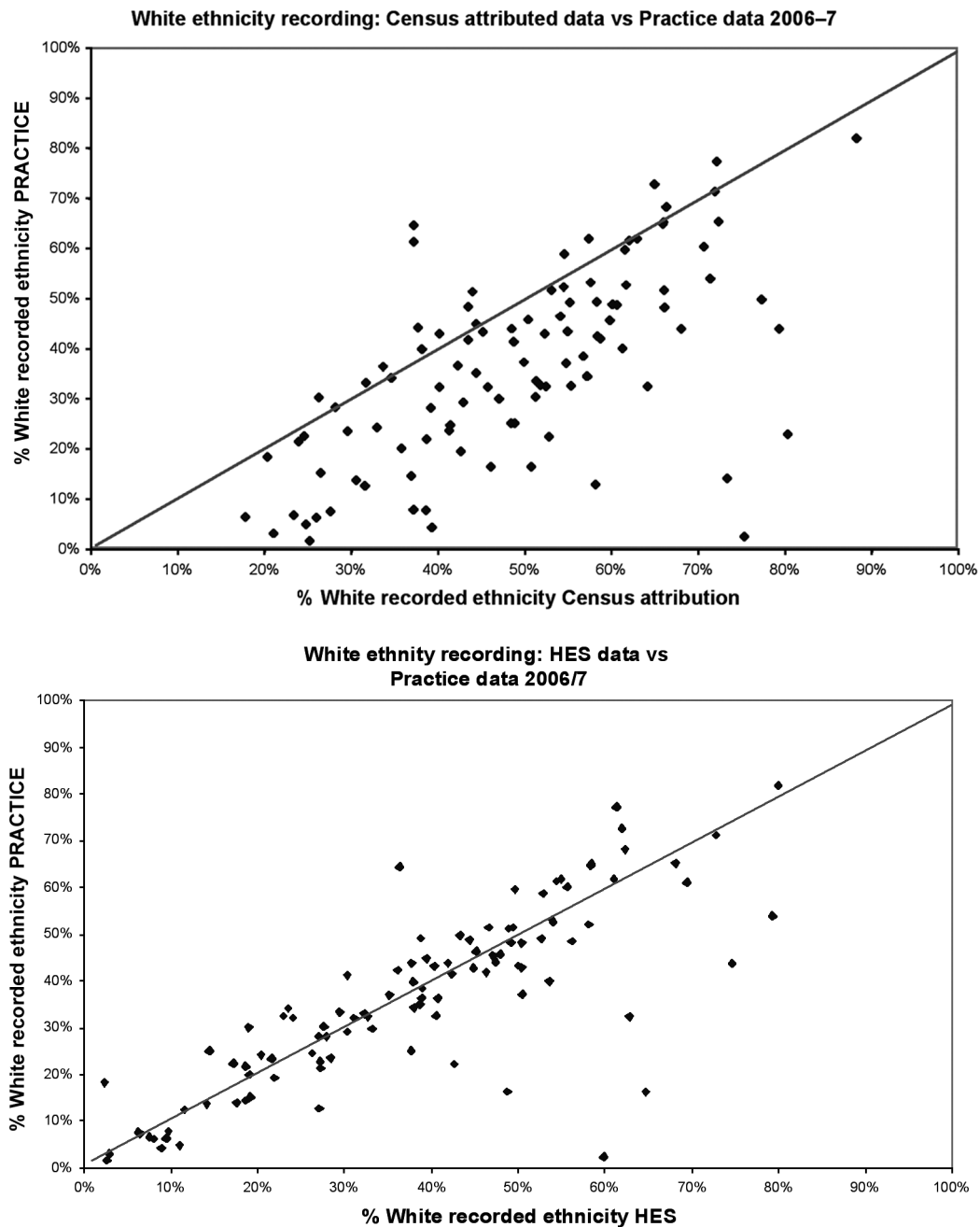


Figure 1a Comparison of census attribution and practice values for % white recorded ethnicity

behave independently of the majority. When applied to general practices in urban areas, there will be selection of practices by sub-groups within the population, depending on ethnicity of the providers, range of services, location, repute and possibly other factors. Hence within a congested urban environment, where there is a choice of practice and where practice boundaries overlap, ethnicity (and probably other factors) will not necessarily be represented by the geographical area from which the practice population is drawn (Figure 3).

Study limitations

All the three methods described rely on self-assessment of ethnicity. East London PCTs have provided training and support for practices to implement this, but there has not been a systematic attempt to review the accuracy at practice level. Similarly there may be variations in performance in the HES data that is captured at entry to hospital.

The census attributed data has further problems. It assumes constancy for demographic data which is not borne out in practice. The census data are only updated

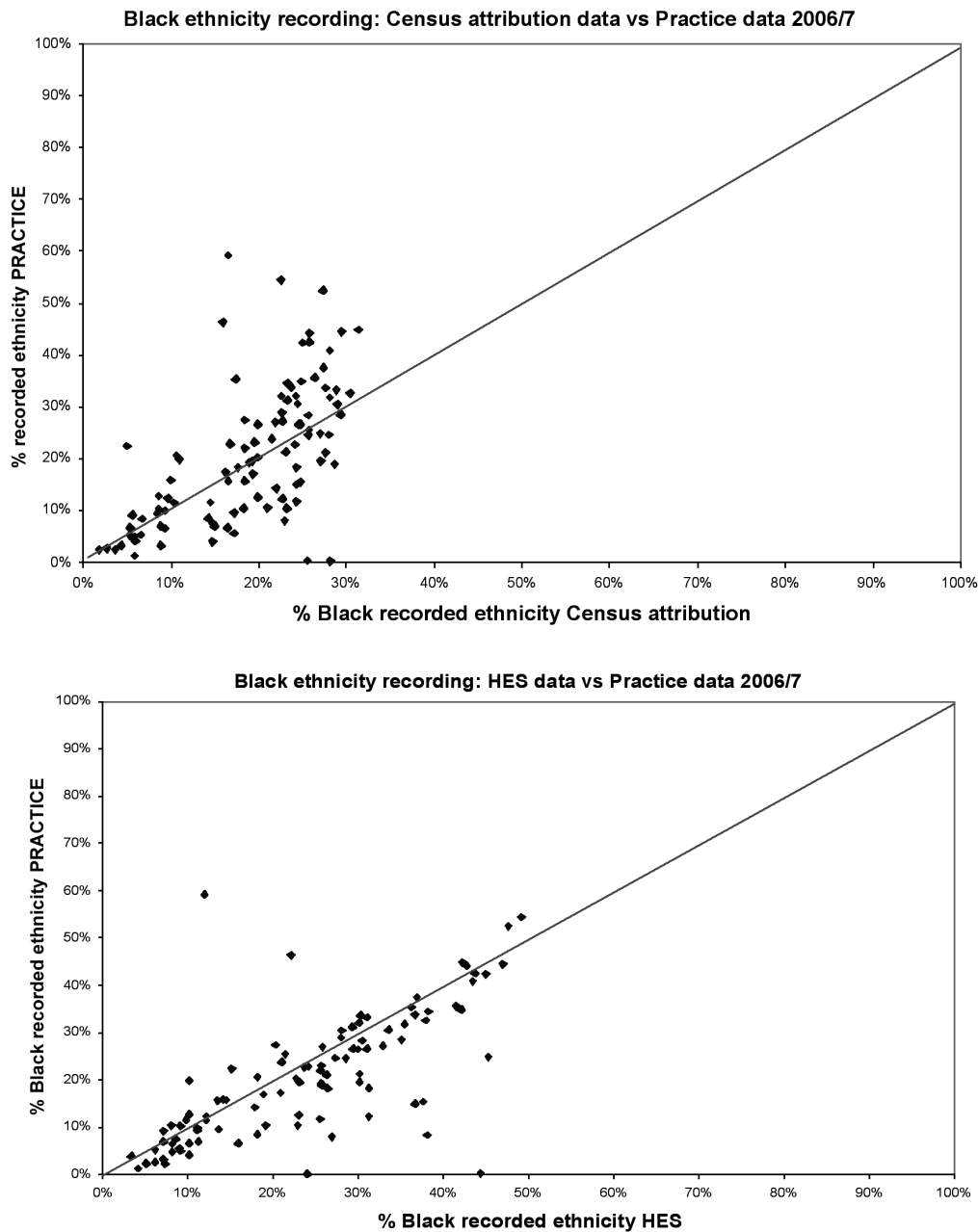


Figure 1b Comparison of census attributed and practice values for % black recorded ethnicity

every ten years, although there is an adjustment at intervals, based on sources such as births and deaths and estimates of domestic and international migration. The ONS figures, although accurate for London overall, conceal over- and undercounting for groups of boroughs (dependant on migration patterns) such that the Greater London Authority has developed a further set of population projections.¹⁹

The HES database has the potential to be influenced by differing access rates (e.g. among south Asian women between 18 and 45 with high fertility rates), although our study did not identify such problems. Calculation of ethnicity by age band will be representative only

where there are high levels of admission. For age groups where admission rates are low, using two years of data may improve the robustness of estimates. The HES data could be used as an additional source for monitoring changes in ethnic populations over time, not only at practice level but for broader geographic areas.

Implications for practice

These findings are significant for a number of reasons.

They illustrate the importance of local incentives as well as national support for the collection of accurate

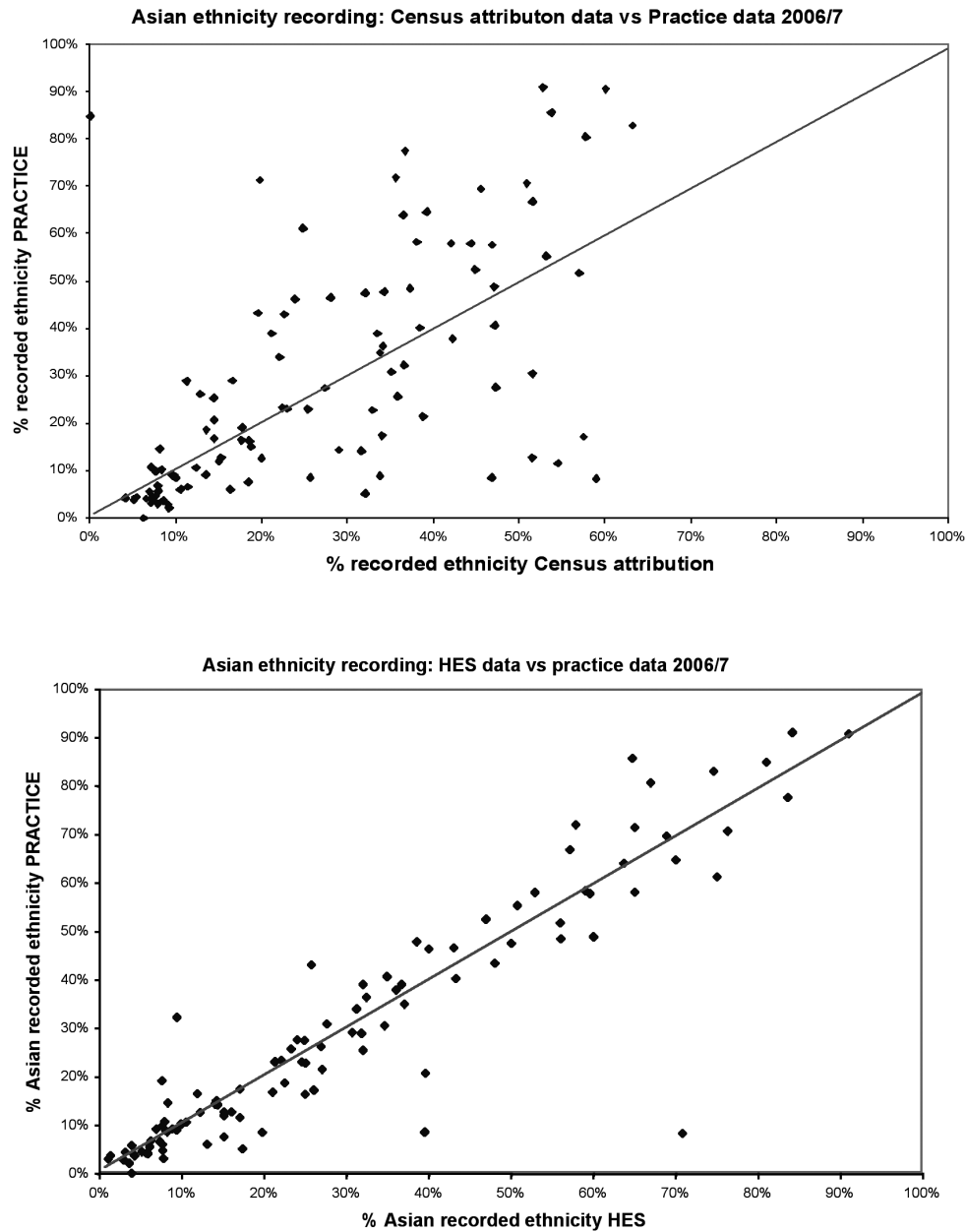


Figure 1c Comparison of ELCIS and practice values for % Asian recorded ethnicity

ethnicity data at practice level. The value of accurate and high recording rates of ethnicity data will be useful both for service monitoring and to support provider initiatives in response to local findings. Population need can be assessed using techniques such as calculating the predicted prevalence of major chronic diseases for a location, and identifying where practices may be under-ascertaining cases. For conditions where the prevalence varies by ethnicity it is essential to have accurate demographic data on practice populations in order to make an accurate prediction for each practice. Public health and practice-based interventions to im-

prove uptake of health programmes, such as immunisation, cardiovascular screening or smoking cessation can also be supported by such data.

Accurate ethnicity data is also important as, in the UK setting, budgets for health care are progressively devolved to practices.^{20,21} The assessment of urban population needs must go beyond the current assumption that the practice population matches the geographic area from which it is drawn. This may become increasingly important as the range of primary care providers grows and the potential for selection of patient populations emerges, and with it the potential

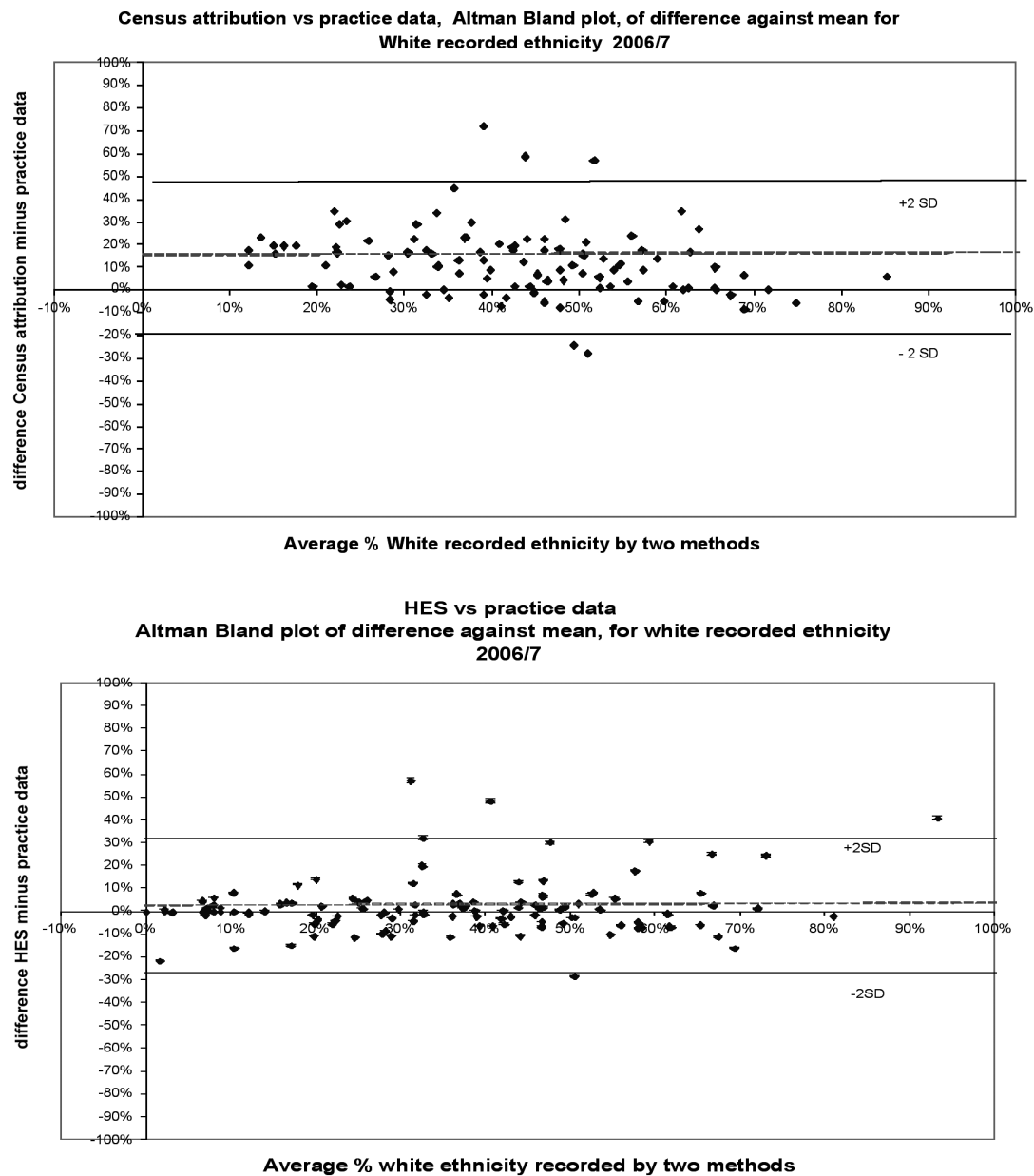


Figure 2a Bland–Altman plots comparing the census attributed data with practice data for white recorded ethnicity

for exacerbating as well as reducing the health disadvantages associated with ethnicity and social deprivation. The use of ethnicity as a service indicator to monitor enrolment at practices, engagement in chronic disease management and other health related programmes will justify the effort required to collect and use this information effectively. These same principles will apply in a range of health systems and settings worldwide.

In the light of these findings we suggest that PCTs in the UK change from using census attributed data to

hospital admission data to estimate ethnic populations when direct practice recording is not available. In 2006 east London practices recorded overall ethnicity for 48% of their populations, and this has risen to over 70% for 2008.¹² Further incentives to promote such recording will reduce the need for indirect methods of measurement. We encourage commissioning organisations throughout the UK to support practice level ethnicity recording in order to develop accurate estimates of disease prevalence and local population needs.

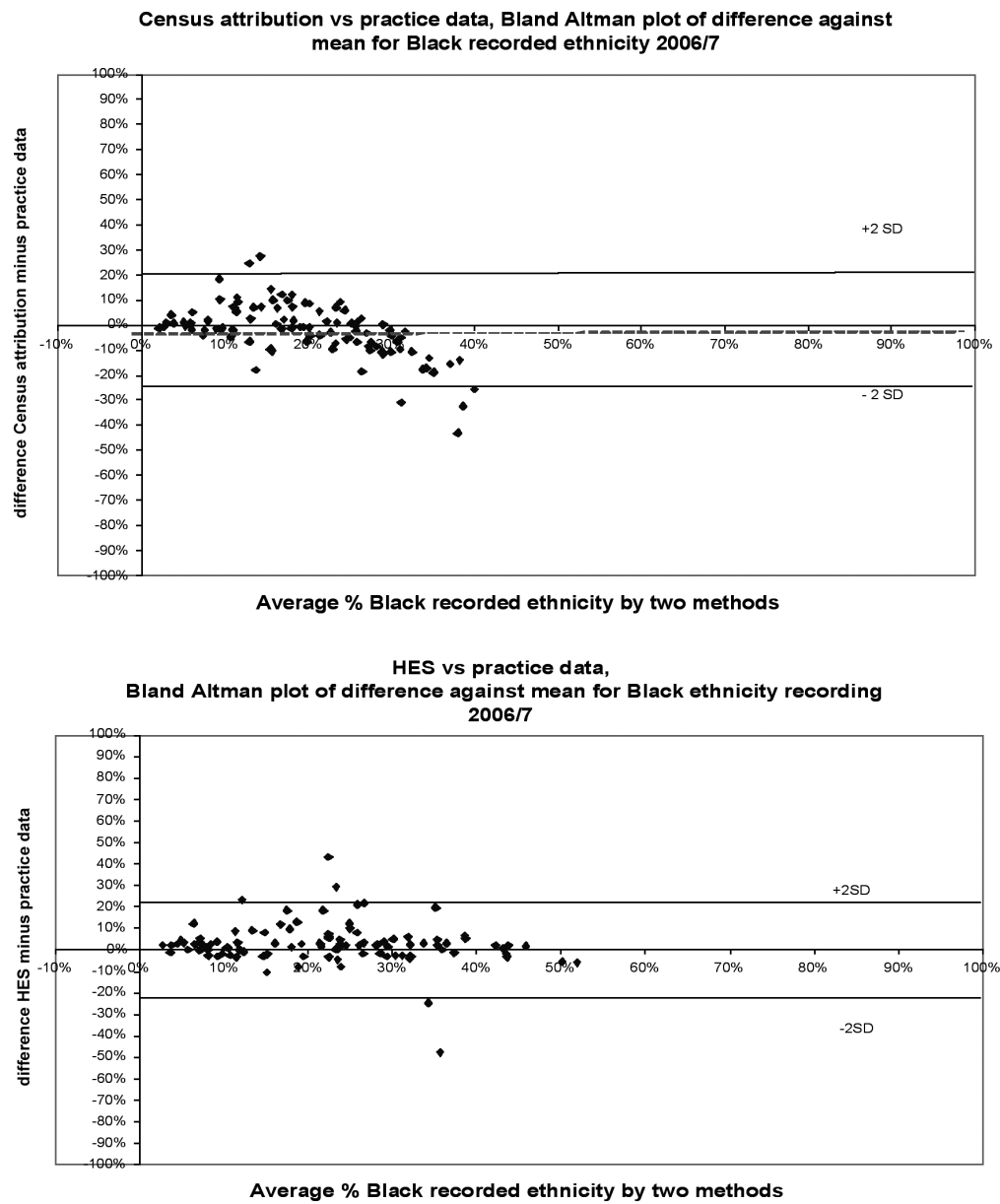


Figure 2b Bland–Altman plots comparing the census attributed data with practice data for black ethnicity

ACKNOWLEDGEMENTS

To east London practices and staff of the CEG who provided the practice audit data. Also to members of the Tower Hamlets PCT ethnicity working group – Anna Livingstone, Mark Caulfield, and Maggie Falshaw – for developing enhanced services to promote practice level ethnicity and language recording.

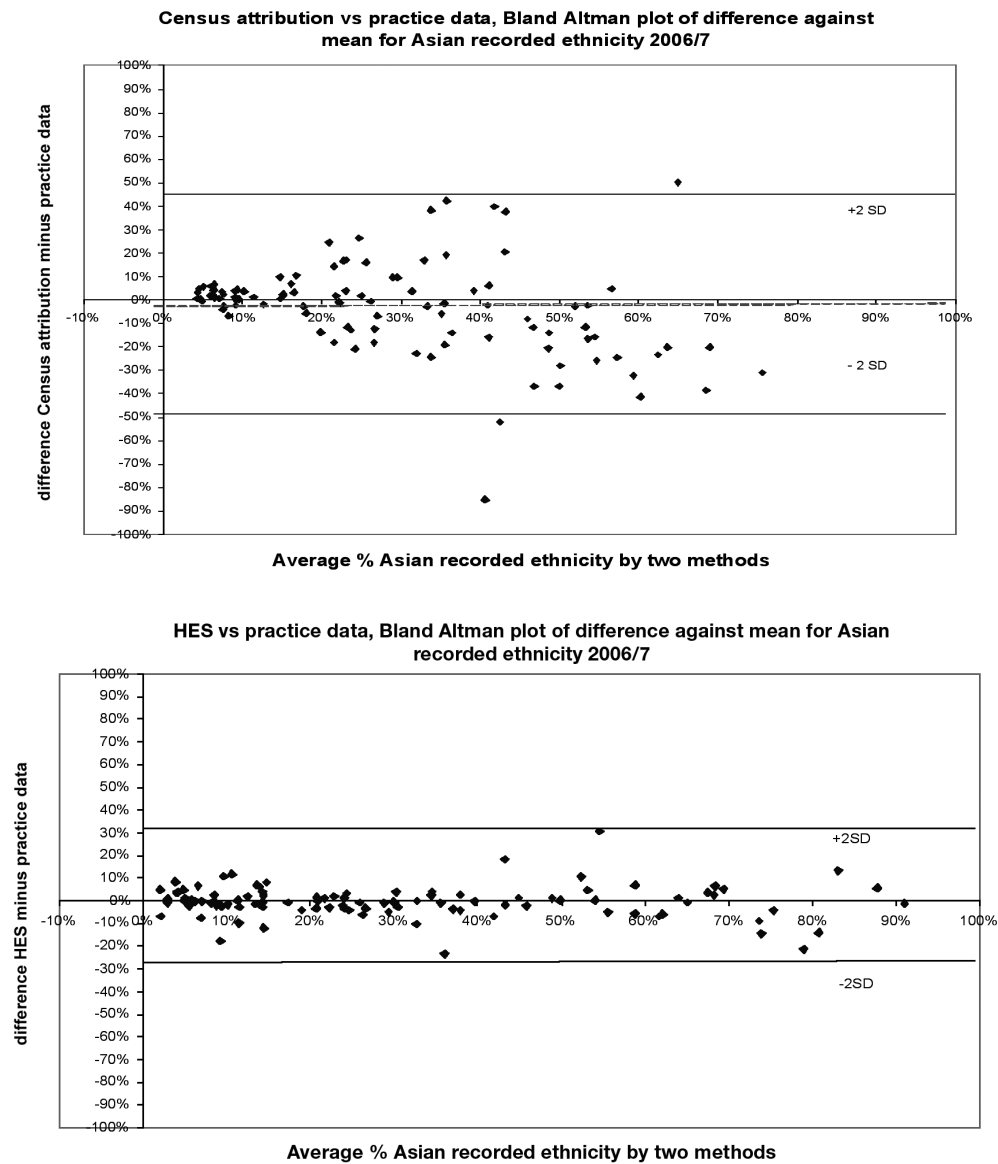


Figure 2c Bland–Altman plots comparing the census attribution data with practice data for Asian recorded ethnicity

Table 2 Summary of mean difference and limits of agreement between different methods of measurement for the main ethnic groups

	HES vs practice data		Census attributed vs practice data	
	Mean difference (95% CI)	Limits of agreement	Mean difference (95% CI)	Limits of agreement
White groups	1.39% (–14.93 to 17.71)	–21.71–24.48	12.49% (–6.15 to 31.14)	–17.65–42.64%
Black groups	3.05 (–11.90 to 18.00)	–16.3–22.42	–1.70 (–17.34 to 13.95)	–22.92–19.52%
Asian groups	0.59% (–14.15 to 15.34)	–18.26–19.45	–2.90% (–24.10 to 18.29)	–41.86–36.05%

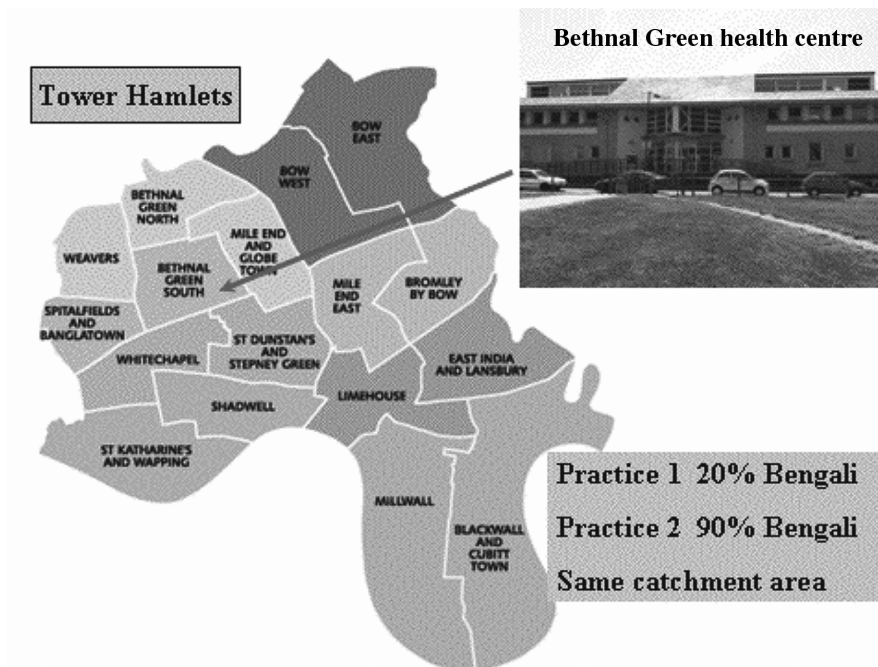


Figure 3 Choice of general practice in Tower Hamlets: an example of the ecological fallacy

REFERENCES

- 1 Department of Health. *Tackling Health Inequalities: a programme for action*. London: HMSO, 2003.
- 2 Fiscella K, Franks P, Gold M and Clancy C. Inequality in quality. Addressing socioeconomic, racial and ethnic disparities in health care. *Journal of the American Medical Association* 2000;283:2579–84.
- 3 Raleigh V. Collection of data on ethnic origin in England. *British Medical Journal* 2008;337:a1107.
- 4 Bhopal R. Is research into ethnicity and health racist, unsound, or important science? *British Medical Journal* 1997;314:1751.
- 5 Senior PA and Bhopal R. Ethnicity as a variable in epidemiological research. *British Medical Journal* 2003;309:327–30.
- 6 McKeigue PM, Shah B and Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in south Asians. *The Lancet* 1991;337:382–6.
- 7 National Institute for Health and Clinical Excellence. *Hypertension: management of hypertension in adults in primary care*. 2006. www.nice.org.uk
- 8 Pringle M and Rothera I. Practicality of recording patient ethnicity in general practice: descriptive intervention study and attitude survey. *British Medical Journal* 1996;312:1080–2.
- 9 Hull SA, Aquino P and Cotter S. Explaining variation in antidepressant prescribing rates in east London: a cross sectional study. *Family Practice* 2005;22:37–42.
- 10 Office for National Statistics. www.statistics.co.uk (accessed 4 May 2008).
- 11 London Health Observatory. www.lho.org.uk/viewResource.aspx?id=11084&uri=http%3a%2f%2fwww.sepho.org.uk%2f (accessed 4 May 2008).
- 12 Clinical Effectiveness Group. www.ihse.qmul.ac.uk/chs/nhs/ceg/index.html (accessed 20 May 2008).
- 13 Robson J, Hull SA, Boomla K *et al*. No ivory towers. *Public Service Review: Health* 2007;10:123–4.
- 14 NHS connecting for Health. MIQUEST. www.connectingforhealth.nhs.uk/systemsandservices/data/miquest (accessed 2 June 2008).
- 15 *Hospital Episode Statistics*. www.hesonline.nhs.uk (accessed February 2008).
- 16 Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;i:307–10.
- 17 MacRae K. Socioeconomic deprivation and health and the ecological fallacy. *British Medical Journal* 1994;309:1478–9.
- 18 Robinson WS. Ecological correlations and the behavior of individuals. *American Sociological Reviews* 1950;15:351–7.
- 19 Greater London Authority Data Management and Analysis Group. *Greater London Authority 2007 Round Demographic Projections*. ISSN 1479–7879. 2008.
- 20 Smith J, Dixon J, Mays N *et al*. Practice based commissioning: applying the research evidence. *British Medical Journal* 2005;331:1397–9.
- 21 Department of Health. *Practice Based Commissioning: promoting clinical engagement*. London: HMSO, 2004.

CONFLICTS OF INTEREST

SAH, KB and JR are general practitioners practising in Tower Hamlets.

ADDRESS FOR CORRESPONDENCE

Sally A Hull
Centre for Health Sciences
Barts and the London Queen Mary's School of Medicine
Abernathy Building
2 Newark Street
London E1 2AT
UK
Email: s.a.hull@qmul.ac.uk

Accepted March 2009

Appendix

These are the mappings we have used:

- **9S1** white to **9i0** British/mixed British
- **9S10** white British to **9i0** British
- **9S11** white Irish to **9i1** Irish
- **9SB5** black Caribbean and white to **9i3** white and black Caribbean
- **9SB6** black African and white to **9i4** white and black African
- **9SB2** Other ethnic, Asian/white origin to **9i5** white and Asian
- **9S6** Indian to **9i7** Indian or British Indian
- **9S8** Bangladeshi to **9i9** Bangladeshi or British Bangladeshi
- **9S7** Pakistani to **9i8** Pakistani or British Pakistani
- **9S2** black Caribbean to **9iB** Caribbean
- **9S3** black African to **9iC** African
- **9S41** black British to **9iD2** black British
- **9S9** Chinese to **9iE** Chinese
- **9SC** Vietnamese to **9iF0** Vietnamese
- **9SD** Ethnic group not given to **9iG** Ethnic category not stated

Appendix Figure 1 Mapping between the 9S and the 9i hierarchies for self-reported practice ethnicity recording

Supplementary data for this paper is available at http://www.radcliffe-oxford.com/journals/J12_Informatics_in_Primary_Care/Supplementary%20Papers.htm