The University
of Manchester

Cathie Marsh Centre for Census and Survey Research

Measuring Ethnic Differences: are surveys up to the job?

CCSR Working Paper 2006-09
Mark Brown
Mark.brown@manchester.ac.uk

The sample survey is valued as a major source of quantitative information about UK society, informing both academic and policy research. Using the investigation of ethnic differences in health as an exemplar, this paper considers the limitations of the sample survey as a research tool for understanding ethnic differences. Specifically it highlights the limitations imposed by sample size, drawing on the authors secondary analysis of the 1999 Health Survey for England (a high profile Government survey specifically designed to facilitate ethnic group analysis). The analysis reveals that statistically reliable comparisons between ethnic groups could only be achieved via a strategy of heavy recoding of age and ethnicity that seriously compromised the theoretical integrity and substantive value of the analysis.

www.ccsr.ac.uk

Measuring ethnic differences: are surveys up to the job?

Introduction: meeting the demand for ethnic group data

The UK is particularly well resourced with social survey data, including an extensive programme of large scale government surveys. In a political climate advocating evidence based social policy, the ability to quantify differences between population sub-groups with a known degree of precision and to generalise findings to the population at large (inference) makes the sample survey a hugely appealing research tool. These qualities help explain the particularly prominent role of surveys in the study of ethnic difference. Against a background of perceived disadvantage and discrimination, more accurate measurement of ethnic differences in a range of outcome measures has long been sought by those wishing to monitor, analyse and address social inequality (Bhrolchain 1990; Ahmad and Shelton 1993; Bulmer 1996; Ahmad 1999; Coleman and Salt 1996; Aspinall 2001). Moreover, the appeal of the survey as a research tool has extended beyond good description, with increased availability of individual level survey data (microdata) enabling application of multivariate techniques to investigate and model the ethnic dimension to many aspects of social, economic and health inequalities.

Prior to the 1991 Census quantitative work on ethnic differences was hampered by a lack of a direct measure of ethnicity in most datasets (the Policy Studies Institutes tradition of ethnic minority surveys being a notable exception: see Smith 1977; Brown, 1984; Modood et al 1997). A reliance on country of birth as a proxy for ethnic group had become increasingly unsatisfactory as UK born second and third generation descendants made up an increasing share of the Black and Minority Ethnic (B&ME) population (around 50 percent for some groups by 1991). The inclusion of a direct question in the 1991 Census was a significant landmark,

not only for providing the first reliable quantification of ethnic minority populations across Britain, but for legitimising the collection of an individual's ethnicity more generally for purposes of social statistics. An ethnic group question quickly became a routine inclusion in all the UKs major social surveys (as it has in administrative datasets for purposes of ethnic monitoring) considerably broadening the scope for measurement of ethnic differences beyond the standard census topics.

With survey data contributing ever more heavily to the evidence base on which social policy is formulated and evaluated, it is important to reflect critically on the nature and extent of it's limitations. This paper highlights one such limitation in specific relation to the measurement of ethnic difference (although it applies generally to the study of population sub-groups). While much of the critique of the survey method in this topic area has centred on the crudeness with which the concept of ethnic identity is operationalised in standardised questionnaires, the focus here is on anther aspect of survey design, the size of the sample. Using an exemplar analysis of the 1999 Health Survey for England (a survey specifically deigned to enable ethnic comparison), this paper will argue that it is the limitations imposed by sample size that represent a more serious and pressing threat to the sensitivity of ethnic analysis than any inherent crudeness of standardised variables.

Using survey data to study ethnic differences in health: a research example

The relationship between ethnicity and health commands a considerable research literature (Nazroo 1997, 1998; Smaje 1995). Fuelled by the numerical increases in the UK B&ME population this has included a growing concern with the ethnic dimension to health service provision, particularly in the major metropolitan areas where B&ME populations are concentrated (Smaje and LeGrand, 1997; Free and McKee 1988). From a policy perspective an important feature of that demographic

increase is the prospect of major increases in the numbers of B&ME elderly, as the main post-war cohorts of first generation migrants (arriving in the 1960s and 1970s) reach retirement age. With the elderly accounting for the vast majority of health care resources in the UK, the question of whether and how health needs may differ by ethnicity has become increasingly salient to health service planning, and is now discussed directly in national policy documents (Department of Health 2001, 2003). Differences may be considered in terms of the overall burden of ill health (do certain ethnic groups experience a higher prevalence of ill health than others?), in the profile of ill health (are certain groups pre-dispoposed to certain types of disability or illness?) and in the nature of the required service provision (are certain forms of provision more or less culturally appropriate for different ethnic groups?).

The research evidence points to differences by all three criteria, for which the theoretical literature offers a number of competing theories (Smaje 1995; Nazroo 1998). These incorporate both material and cultural explanations for ethnic differences in the nature and level of need, both of which may be especially pertinent to the elderly population.

Concerning material explanations, a plethora of local studies over recent decades confirms the B&ME elderly as a particularly vulnerable social group (Norman 1985; Donaldson 1986; Rich et al 1996; Evandrou 2000; Rich et al 2001), and tending to adhere most strongly to traditional cultural values and lifestyles (Modood et al. 1994; Jewson et al. 2003) Crucially, such studies are challenging the previously held notion that the care needs of the elderly (particularly those of South Asian origin) are being met within their own families and communities (Murray and Brown 1998; Department of Health 1998; Karn et al. 1999). Low uptake of health and related social services is now increasingly recognised as a problem of accessible and appropriate service delivery rather than low needs in the community

(Smaje 1995; Smaje and Le Grand 1997; Free and McKee 1998; Shah 1998; Walker and Ahmad 1994; Ahmed and Walker 1997) However, most of the recent empirical evidence comes from small and mainly qualitative studies. There remains a relative scarcity of robust quantitative evidence on the health of the B&ME elderly, particularly evidence that can be generalised and allow reliable ethnic group comparison at national level. Progress in understanding the causal mechanisms behind health differences, as well as the formulation of a strategic policy response, would certainly benefit from a more scientific measurement of the differences we are seeking to explain.

Against this background, the 1999 version of the Health Survey of England (HSE), with a boosted B&ME sample, came as a welcome addition for quantitative researchers in this field. However, it is the contention of this paper that the survey's potential to describe and explain ethnic group differences is severely limited by a sample size that prevents all but rudimentary analysis by age. This argument is developed using evidence from the authors' own secondary analysis of the HSE data, which entailed the relatively straightforward objective of generating national estimates of various health measures by age for different ethnic groups.

While a number of different health measures were considered in this work, analysis reported in this paper focuses on a measure of psychological wellbeing. Mental health is an area where ethnic differences (particularly in old age) appear to be most profound, in terms of the level and nature of the underlying morbidity burden (Abas 1996; McCracken et al. 1997; Rait and Burns 1997; Department of Health 2003), and especially in terms of a widely perceived failing of services to meet needs (Department of Health, 2003). Given the general absence of reliable statistics in this area, robust estimates of psychological morbidity by ethnic group is precisely the kind of output the 1999 HSE (with a boosted ethnic minority sample

and a rare measurement of self-reported mental health (the General Health Questionnaire (12) or GHQ12[1]), might be expected to deliver.

The analysis

A useful starting point for the analysis is the very detailed official report of the 1999 HSE (Erens et al. 2001). The analysis of the GHQ(12) question (Calderwood and Tait 2001), in common with the rest of the report, relies heavily on the use of age-standardised risk ratios as the basis for measuring ethnic difference. These reveal notable differences between groups in the level of psychiatric morbidity (defined here by a GHQ12 score of 4 or more, subsequently refered to as a 'High GHQ12 score'), with Bangladeshis significantly more likely than the general population to report a high score and the Chinese significantly less likely (Table 1).

TABLE 1 ABOUT HERE

Comparisons based on a single age-standardised indices, by definition, provide no insight into whether the observed differences between ethnic groups (or lack of difference) apply equally across all ages. And yet, sensitivity to differences in the age patterning of ill health must be considered crucial for an effective description of ethnic differences in ill health, and particularly an accurate measure of individual risk within ethnic groups. Moreover, from a policy perspective, such a description is likely to be an essential requirement for the accurate assessment and measurement of service needs in any population defined at national or local level. In short, we

---

[1] The GHQ (12) is a set of 12 questions designed to detect psychiatric morbidity. The questions cover general levels of happiness, depression, anxiety and sleep disturbance. A derived score of 4 or more (referred to as a 'high GHQ12 score') is used in the report to identify informants with a possible psychiatric disorder. Although validated for the general population, the GHQ12 has not been specifically validated for use with the B&ME population requiring caution in the interpretation of results (Calderwood and Tait, 2001)

need to know whether the disproportionately high psychological morbidity reported by Bangladeshis is consistent over the life course or concentrated among certain age groups. However, while observed differences in the age pattern of high GHQ12 score are alluded to in the main text of the report (and shown in a linked table to the on-line report) these are not the basis for further discussion or analysis:

In the general population, the proportion with a high GHQ12 score did not vary much with age, but among men of South Asian origin and Black Caribbean men, it appeared to increase with increasing age... However, the samples of each sex in the oldest age group (55 and over) were small, and should be treated with caution.

(Calderwood and Tait 2001)

The different pattern of ill health by age between ethnic groups is a finding of major interest, particularly in terms of the research focus on the elderly, but with more general implications for the way the risk ratios in Table 1 are interpreted (e.g. the fact that young South Asians do not experience disproportionately high GHQ12 scores is masked in the single indices). The final reports inability to highlight this finding, or investigate it further, because of insufficient sample size suggests a real shortcoming of the survey design. The extent and nature of the limitations of sample size were consequently investigated more explicitly in a secondary analysis of the survey microdata (made available via the Economic and Social Data Service (ESDS)[2].

---

[2] See http://www.esds.ac.uk/government/

The Analysis

Figure 1 shows the prevalence of a 'high' GHQ12 score by age and ethnic group. As a starting point the analysis uses four categories of age rather than the three used by default in the official survey report, employing age 65 rather than 55 as the cut off to define the elderly population. While no clear age patterning is evident in the general population, Figure 1 confirms some steep increases in the prevalence of high GHQ12 scores by age for the South Asian groups (steeper than was apparent in the original published table with its fewer categories of age). Comparing groups within age bands it is evident that most of the overall difference in risk is down to differences among the elderly population. For the population below the age of 35 there is little variation across ethnic groups. Indeed Indians even report a slightly lower prevalence with a high GHQ12 score than observed for the general population. At the other extreme, among those aged 65 and over, Bangladeshi women in the sample experience almost three times the prevalence observed in the general population.

FIGURE 1 ABOUT HERE

The patterns revealed in Figure 1 are of considerable substantive interest but their research value is clearly compromised by the relatively wide confidence intervals around most of the estimates. This reflects relatively small sample sizes which result in some very large standard errors. For some sub-groups (Bangladeshi and Pakistani men and women aged 65 and over) the estimates are based on less than 30 cases, too few to enable calculation of meaningful confidence intervals.

The small numbers make it very hard to report any of the observed differences as statistically significant, even where these are large in absolute terms.

For example, while the raw estimates suggest substantial differences between Indian, Pakistani and Indians at all age groups, in no case is this a statistically significant difference. For Pakistani estimates in particular, confidence intervals are so wide that they even prevent the confirmation of a statistically significant difference from the general population. The same problem arises when comparing age groups within ethnic groups. Thus despite an impression of strong age patterning to health in the south Asian groups, the differences by age are only statistically significant for Bangladeshis. It is often overlooked that failure to attain statistical significance does not signify that an observed association in the sample is in any sense 'false' or unimportant in the wider population, simply that we lack the sample size to confirm it as 'real' (at an arbitrarily set level of probability). Moreover, the probability that the difference between two estimates is 'real' may still be higher than the probability that the difference arose by chance.

Despite this, and the fact that the findings reported in Figure 1 (if verified) carry implications for both policy and theoretical understanding, in terms of the normal criteria applied to empirical evidence of this type, their lack of statistical significance means they are unlikely to be highlighted or even reported in research outputs. Hence the very cautious treatment in the main HSE report. The situation is one frequently encountered when undertaking secondary analysis of sample surveys, and highlights a fundamental problem where the research questions being investigated have not directly informed the design of the survey (a central principle in all the classic works on survey design).

Where, as for the output reported in Figure 1, sample size is insufficient to support the level of disaggregation, researchers typically adopt a strategy of variable recoding to boost cell counts (and so reduce standard errors). In this example, the options are a re-coding of age, ethnicity or both. Table 2 reports the impact of

various recoding schemes on the ethnic group comparison of 'High' GHQ12 scores (for clarity, this is shown for South Asian groups only and with a focus on the elderly). This includes a recoding of the top age category to include those aged 55+ (rather than 65+) and two additional recodes of ethnic group (Pakistani and Bangladeshi combined and a single South Asian category). Specifically, Table 2 shows how the recoding strategies effect the measured difference in 'High' GHQ12 score estimates, i) between ethnic groups by age, and ii) by age within ethnic groups, and whether or not the difference is statistically significant.

TABLE 2 ABOUT HERE

While the underlying crosstabulation remains ostensibly the same (High GHQ12 score by age by ethnic group), revising variable categories is seen to have a fundamental impact on the measurement of difference. As a generalisation, the aggregation of variable categories has the effect of reducing the amount of reported variation in the population for that variable. In this example, where the underlying pattern is for the share with a 'High' GHQ12 score to increase with age, revising the elderly category to include those aged 55 and over (the cut-off used in the HSE report) rather than 65 and over (as used in the analysis shown in Figure 1) reduces the measured differences between young and old within ethnic groups, and that between the 'elderly' of different ethnic groups, in almost all cases. For example, for 'elderly' men, the difference between the Indian and general population is almost halved when the revised classification of age is used. However, this is not true for Indian women, for whom the percent with a 'High' GHQ12 score is actually higher among those in the sample aged 55 to 65 than those aged 65 and over.

The reported pattern of difference is further affected by the revision of ethnic group categories. By adopting the often used practice of grouping Pakistani and Bangladeshi together in situations of small sample size, observed differences in the age pattern and level of High GHQ12 score between the two groups (admittedly based on too few cases to consider reliable) are subsumed into an average that may represent neither accurately. This problem becomes even more extreme if we include Indians to form a single South Asian category. This ensures estimates for the elderly that are more robust in statistical terms, but which now mask major within-group heterogeneity. Thus the fact that elderly Indians in the sample report a lower prevalence of High GHQ12 scores than the general population is lost in a single South Asian estimate, that actually suggests the opposite.

Heavy recoding in this way has clearly reduced the sensitivity of the analysis to variation in the survey population. Of equal concern, and independent of underlying measurements, it has resulted in variable categories that are conceptually and theoretically unsound. First, while determining an age cut-off to define the elderly is inherently subjective, the use of a single 'age 55 and over' category to define the elderly in a study of health lacks any obvious theoretical, policy or empirical justification. Increases in life expectancy and changing working patterns have resulted in a growing and increasingly diverse post-retirement age population. This has led to growing recognition of the need (for research and policy purposes) to distinguish between 'young' elderly (typically active and in good health) and 'frail' elderly (those in their 80s and 90s). It is self evident that health status is one aspect where such a distinction is likely to be particularly important, especially where (as in the current analysis) there is evidence that the relationship between age and the measure of interest varies by ethnic group. A further reason for avoiding a single broad and open ended age category when conducting ethnic

group comparisons is the potentially confounding impact of differences between ethnic groups in the distribution of age (the mean age of the B&ME elderly tending to be younger than in the general population, because of the relatively small numbers of very old people).

The use of 'age 55 and over' to define older people in all age-disaggregated findings presented in the official 1999 HSE report stands in notable contrast to the age detail available in comparable output from the 2001 Census. Based on 100 percent population data, census tables crosstabulating health by ethnicity distinguish three separate categories among the population aged 55 and over (55-64; 65-74 and 75+). Although the health measures in the census are restricted to general health and limiting long term illness (and so are not directly comparable with the HSE analysis reported here) the differences recorded between these three age groups (discussed in more detail below) illustrate the potential loss of sensitivity to age differences implied by the use of a single age 55 and over category in the HSE report.

Turning to the recoding of ethnicity, the relatively small size of B&ME populations, and their generally youthful age structure makes some aggregation of ethnic categories the norm in survey analysis, especially where there is a requirement for age-disaggregation. However, many have questioned the validity of these groupings and of the underlying classifications on which they are based (Ballard 1994; Modood et al 1997; Aspinall 2001; Nazroo 1998). As an illustration, in a local study of Blackburns Asian minority, Robinson (1975) used differences in regional origin, religion and language to identify seventeen cultural sub-groups. Seen in this context, the three categories of South Asian identity used in the standard ethnicity variable (Indian, Pakistani and Bangladeshi) match only poorly to underlying cultural diversity. The further loss of definitional integrity evident in the

creation of a Pakistani/Bangladeshi or single South Asian category must cast further doubt over the interpretation of 'ethnic group' estimates, and particularly of differences between them in a comparative analysis. Given the lack of an explicit or consistent theoretical framework for the definition of ethnic categories it is important to question whether we are actually measuring an ethnic group difference at all? As agued by Nazroo, the use of such culturally diverse and un-theorised groupings makes it very hard to directly assess 'cultural' explanations for any observed differences between groups (e.g. explaining health differences in terms of lifestyle characteristics such as diet, exercise, or the utilisation of health care services) (Nazroo, 1998). It also reduces the utility of ethnic group statistics for those involved in service provision, especially at the local level, a point stressed by Robinson in justifying his use of the 17 point ethnic classification in his Blackburn study (Robinson, 1975). The same point is made by Ahmad and Shelton (1993), adopting the perspective of a health authority trying to improve its service delivery by making it more sensitive to cultural differences in need (the following specifically concerns the diverse nature of dietary and language needs in a local community):

> To offer appropriate diet these categories become meaningless. 'Indian'….'Black Caribbean' 'Black African' for example tell us nothing about diet habits. An Indian may be a Punjabi, Bengali or Gujarati; Muslim, Hindu, Sikh or Christian; vegetarian or meat eater; and among meat eaters, requiring (or wanting) halal meat or non-halal meat; rice eater or chapati eater. If the same authority wishes to improve its interpreting services, then the category 'Indian' tells it nothing about the mix of languages spoken (for example, Punjabi, Urdu, Gujarati, Hindi, Bengali).
> (Ahmad and Shelton 1993)

The analyses of the 1999 HSE highlights a difficult trade-off between the need to retain variable detail and the need to produce results that are robust and statistically significant. That trade-off is determined in large part by the available sample size. It is the central argument of this paper that the sample size of the 1999 HSE (and of social surveys more generally) is too small to allow more than a rudimentary analysis of ethnic differences. Efforts to undertake a simple age-disaggregated comparison of mental health required heavy recoding on both age and ethnic group to generate estimates that were statistically robust. That procedure resulted in a serious loss of sensitivity to observed variation in the underlying data and a set of estimates for groups that lacked theoretical integrity, making their interpretation, for academic or policy maker, extremely difficult. Thus, while many critics of the survey method focus on a perceived lack of sensitivity to differentiation inherent in the process of measurement (the forcing of a concept as complex as ethnic identity into 10 or 16 standard categories), in this example sample size imposed a far more severe restriction on the researchers ability to represent ethnic (and age related) diversity – with a resulting level of detail in the analysis that fell well short of the potential in the underlying variables available on the dataset. Unfortunately, we are unable to quantify the amount of 'lost detail' resulting from the recoding schemes in Table 2, for the very reason that the estimates prior to recoding lack statistical robustness (we do not know how much of the underlying pattern in the survey data is distorted by sampling error). In this context, a second research example drawn from the 2001 Census (100 percent tables), provides a more robust illustration of the extent to which recoding schemes are liable to mask important differences in the population.

Ethnic differences in health using the UK Census

In terms of variable detail, the UK Census offers less scope than the 1999 HSE to measure ethnic differences in health, being restricted to two measures of self reported health: long term limiting illness and (from 2001) general health. Moreover, at least in terms of the 100 percent Census data, the researcher is limited by what is made available in the form of aggregate tables. These lack the flexibility afforded by micro-data including the potential to undertake multivariate analysis. However, in terms of the task of basic description, use of population tables have obvious advantages over sample based estimates, assuming they contain the required variable specification. In this case we are interested in estimates of 'not good' general health by age by ethnic group. This is the subject of a commissioned table made available by the Office of National Statistics (ONS), which provides the basis for Figure 2. This shows, for males and females separately, the percentage point difference in the prevalence of 'not good' general health between South Asian ethnic groups and the general population, for different age groups. In contrast to the single elderly category of the survey analysis reported earlier, the Census table differentiates the population aged 55 and over into three categories. It also allows us to work with the full 16 point ethnic group classification of the 2001 census (though only South Asians are shown in Figure 3).

FIGURE 2 ABOUT HERE

Thus, although the Census questionnaire is much more limited in scope and detail than the HSE, the Census ultimately allows a more detailed measurement of the three way association between health, ethnicity and age. Also, because we are

working with 100 percent population data, we may assume that differences revealed in Figure 2, even where small in absolute terms, are 'real'. Since the dependent variable (general health) is different from that used in the 1999 HSE analysis (GHQ12 high score), it is not possible to make any direct comparisons with the survey analysis[3]. However, the census data reveals some striking variation in ill health (by age and ethnic group), that would be masked if subject to the same recoding strategies employed in the HSE analysis.

Figure 2 echoes the HSE analysis of GHQ12 in revealing that the health disadvantage of South Asians (relative to the general population) in general health relates to the middle and older age groups, and not the young. But it also reveals notable differences between South Asian populations and by sex in the age pattern and extent of disadvantage after age 60. For Pakistanis and particularly Bangladeshis, there is a striking reduction in relative disadvantage among the very old (over age 75), which in the case of Bangladeshi women, actually appears to become a slight relative advantage. It is clear that some of the differences observed here may be artefactual. Thus the anomaly of the over 75 age group may partly reflect a lower mean age for South Asians in this last open ended group. More generally, comparisons of self reported health may be affected by differences in the way survey questions are interpreted and answered according to age and ethnicity, rather than in underlying physical or mental health. The search for an explanation for these contrasting patterns lies beyond the scope of this paper. The important point to be made here is that the process of secondary analysis starts with good description. This paper argues that the sample sizes in our major surveys are currently inadequate for that task, at least in respect to the description of ethnic group differences.

---

[3] In fact analysis of the HSE (not shown), which includes a general health question, suggests a strong correlation between the two measures.

Discussion: Curse of the Small sample

The sample survey is valued as a major source of quantitative information about UK society, informing both academic and policy research. However, in considering the contribution of sample surveys to the understanding of ethnic differences, evidence presented here has raised questions about their 'fitness for purpose'. Specifically, the paper observes an inability to support more than rudimentary description of ethnic differences by age, even where surveys have a specific design remit for ethnic comparison. Crucially the root cause for this failing is not the inherent lack of variable detail in the datasets, but an inadequate sample size. As an illustrative example, the paper has evaluated the potential of the 1999 HSE to inform on a topic of growing interest - the health of B&ME elderly. Statistically reliable comparisons could only be achieved via a strategy of heavy recoding of age and ethnic groups that seriously compromised the theoretical integrity and substantive value of the analysis.

The limitations described for the 1999 HSE analysis are considerably more severe when using national surveys with sampling designs that apply selection probabilities proportional to population size (the majority of government surveys). While such surveys typically include an ethnic group variable sample sizes leave little scope for disagregation. While techniques such as regression cope more effectively with small samples than crosstabulation, and are frequently employed on such datasets to investigate ethnic effects, the failure to support basic description is an important limitation of their contribution to a better understanding of ethnic differences in UK society. It is a measure both of the extent of this limitation (and of the strength of interest in ethnic group analysis) that boosted ethnic samples are now an increasingly common modification to survey design (routinely used in some

major Government surveys including the British Crime Survey). However, as this paper highlights the 'boost' to samples may still fall well short of the numbers required to support statistically significant findings at the level of disaggregation required by researchers – which for ethnic group comparisons, almost always needs to include disagregation by age and sex as a minimum. However, in drawing lessons from the reported HSE analysis, it should be acknowledged that the challenge of attaining workable sample sizes for the B&ME elderly is a particularly difficult one due simply to their relative scarcity in the B&ME population from which the sample is drawn.

In focussing on sample size as a constraint on the utility of surveys for ethnicity research there is no intention to gloss over the shortcomings inherent in other aspects of research design. This includes reservations expressed by many commentators over the validity of the ethnicity variable itself (Ahmad 1999). Moreover, in the reported HSE analysis, there are obvious question concerns the validity of the GHQ12 variable as a measure of psychological morbidity, particularly in the context of ethnic group comparisons. Nevertheless, from the evidence presented here it is clear that a main culprit behind the tendency towards inadequate differentiation in ethnic group analysis is an inadequate sample size as much as underlying shortcomings in the variable definitions. Indeed, by incorporating additional variables on religion and language, surveys such as the HSE have the potential to support a far more sophisticated classification of cultural identity than is ever incorporated in analysis, because of the restrictions of sample size.

As already observed, the problem here is not only the general masking of population heterogeneity that necessary recoding implies, but the fact that the collapsing of groups may substantially reduce the descriptive and explanatory power

of the analysis in a way that leads to over-simplistic and even erroneous interpretation of the measured associations of interest.

Finally, what recommendations arise from this critique of the sample survey? First, it is important to emphasise that sample size is an aspect of survey design and as such the limitations identified here relate to a practical design matter rather than any fundamental rejection of the survey method itself. However, the logical design response to increase the sample size of surveys has major implications for cost. Where budgets can be increased, investment in bigger samples promises real pay back in terms of the breadth and depth of analysis possible. The utility of the Government surveys to inform understanding of ethnic differences is certainly enhanced by the practice of boosting ethnic samples, but it is apparent that samples would need to be considerably bigger still in order to utilise anything like the full potential from the underlying data.

On the basis that surveys must be delivered to finite budgets, sample size must be considered in tandem with other aspects of survey design. In crude terms, we may see all survey designs involving a trade-off between sample size on the one hand and unit costs per interview on the other. A key component of the latter is obviously the length and depth of survey questionnaires, and other things being equal a shorter quicker survey instrument may facilitate a bigger sample for the same cost. While the depth of information collected in specialist Government surveys like the HSE is highly valued, it is apparent that small numbers are effectively making much of the rich variable detail in these surveys redundant, particularly where analysis seeks to make sub-group comparisons. In many cases variables require heavy recoding even to support the most basic crosstabulations. In terms of the analysis reported here, it is significant that with a far more restricted range of variables, and without the flexibility of micro-data, the census was able to

offer a more detailed breakdown of ethnic differences by age than the specialist survey considered here.

In weighing up the respective advantages of Census tables versus the inherent flexibility and richness of sample survey data, the Sample of Anonymised Records (SARs) from the Census stands out as a unique 'alternative' data source for the research of ethnic differences. An innovation of the 1991 Census, the SARs consist of files of individual level census records. In 1991 they included two files: a one percent sample of households and a two percent sample of individuals. In many ways akin to survey data, the key distinction that marks the SARs apart from the wealth of other sample data in the UK is the size of the sample. As a specific illustration of the scope it offered for ethnic group analysis, the 1991 Individual SAR included more around 17,000 individuals of Indian descent, compared to just under 1,700 in the 1999 HSE ethnic boost sample. The bigger sample in the SARs has enabled researchers to conduct a far more rigorous analysis of ethnic differences than previously possible with other survey data, including age-disaggregation and the retaining of detail on the ethnic classification. As a result the 1991 SARs made a considerable contribution to the better description and understanding of ethnic differences in the UK (Li 2004).

Following the success of the 1991 SARs, similar files have been made available from the 2001 Census. The sample size for the individual SAR has been increased to three percent and the file includes the full 16 point ethnic group classification (the combination of the bigger sampling fraction and population increase since 1991 means the sample now includes over 30,000 Indian cases). Although limited in its range and depth of variables (relative to surveys like the HSE) the 2001 SARS thus represent an exciting data source for researchers seeking

better quantitative description and understanding of ethnic group differences in UK society.

As a final recommendation, it may be helpful to adopt a more openly reflective position regarding statistical significance and its impact in shaping data analysis and the reporting of research. This is in no way to advocate a less rigorous reporting of statistical output, but to caution against the tendency for the constraints of sample size to override a proper regard for theory and the complexity of the underlying distribution of survey data. As illustrated in this paper, the price for attaining statistically significant findings may be a substantial masking of underlying diversity. This may lead to a situation where reported differences, whilst statistically significant, are hard to interpret because the aggregated groupings lack theoretical or substantive integrity or meaning.

Adopting this perspective, the differential age patterning to High GHQ12 scores revealed in the 1999 HSE analysis (Figure 1) is arguably worthy of greater attention than given in the official survey report, where age differences are entirely masked in the single age-standardised risk ratios that make up the headline findings. While lacking the precision to be presented as hard evidence, at the very least they should be presented and discussed as a basis for the development of hypotheses and further research, and a more explicit reminder of the need for bigger samples.

References

ABAS, M. 1996 'Depression and anxiety among older Caribbean people in the UK' *International Journal of Geriatric Psychiatry*, vol11, pp. 377-82.

AHMAD, W. 1999 'Ethnic Statistics: better than nothing or worse than nothing?' in Dorling D and Simpson. S (eds), *Statistics in society: the arithmetic of politics,* London, Arnold: pp.124-32.

AHMAD, W. and SHELTON, T. 1993 '*Race' and Statistics. Social Research: Philosophy, Politics and Practice*, Hammersley, London, Sage

AHMAD, W and WALKER R. 1997 'Asian Older people: housing, health and access to services' *Ageing and Society*

ASPINALL, P 2001 'Operationalising the collection of ethnicity data in studies of the sociology of health and illness' *Sociology of Health and Illness*, Vol 22, 6 pp.829-862

BALLARD R. 1994 *Desh Pardesh, the South Asian Presence in Britain*. Hurst and Co

BHROLCHAIN, N. 1990 'The ethnic question for the 1991 Census: background and issues' *Ethnic and Racial Studies*, vol 13, pp.543-67.

BULMER, M. 1996 'The ethnic group question in the 1991 Census of population', in Coleman D and Salt J *Ethnicity in the 1991 Census, Volume 1: Demographic Characteristics of the Ethnic Minority Populations*, London, HMSO.

CALDERWOOD, L. and TAIT, C. 2001 'Self-reported health and psychosocial well-being' in Erens, B, Primatesta. P and Prior, G. (eds) *Health Survey for England: The Health of Minority Ethnic Groups Volume 1: Findings*, London, The Stationery Office.

COLEMAN, D. and SALT J. 1996 'The ethnic group question in the 1991 Census: a new landmark in British social statistics', in Coleman D and Salt J *Ethnicity in the 1991 Census, Volume 1: Demographic Characteristics of the Ethnic Minority Populations*, London, HMSO, pp.1-33.

DEPARTMENT OF HEALTH 1998 *They Look After Their Own Don't They?*, Social Services Inspectorate.

DEPARTMENT OF HEALTH 2001 *National Service Framework for Older People.*

DEPARTMENT OF HEALTH 2003. *Inside Outside: Improving Mental Health Services for Black and Minority Ethnic Communities in England,* (National Institute for Mental Health in England. (DoH)

DONALDSON, L. 1986 Health and social status of elderly Asians: a community survey', *British Medical Journal* (293): 1079-82.

ERENS, B. and PRIMATESTA P 2001. *Health Survey for England: The Health of Minority Ethnic Groups Volume 1: Findings*, London, The Stationery Office.

EVANDROU, M. 2000 'Social Inequalities in later life: the socio-economic position of older people from ethnic minority groups in Britain', *Population Trends,* vol 101, pp.11-17.

FREE, C. and MCKEE, M 1998 'Meeting the needs of black and minority ethnic groups (The New NHS: Commentaries on the White paper. From Specialist Services to Special Groups. (United Kingdom, National Health Service)' *British Medical Journal* 316(7128), 380(1).

JEWSON, N., JEFFERS, S and KALRA, V. 2003. *Family Care, Respite services and Asian Communities in Leicester*, University of Leicester.

JOHNSON, M. 1993. 'Health and social services (in United Kingdom; report)' *New Community* 19(4): 675(12).

KARN V, MIAN S, BROWN M, and DALE, A. 1999. *Tradition Change and Diversity: understanding the housing needs of minority ethnic groups in Manchester*, The Housing Corporation.

LI, Y. 2004, 'Samples of Anonymized Records (SARs) from the UK Censuses' *Sociology*, 38, 3, pp.553-572.

MCCRACKEN, C, BONEHAM M, COPELAND J, WILLIAMS K, WILSON K, SCOTT A, MCKIBBIN, P and CLEAVE, N. 1997 'Prevalence of dementia and depression among elderly people in Black and ethnic minorities' *British Journal of Psychiatry* 171: pp.269-273.

MODOOD, T, BEISHON, S and VIRDEE S. 1994. *Changing Ethnic Identities*, London, Policy Studies Institute

MODOOD, T, BERTHOUD, R, LAKEY, J, NAZROO, J, SMITH, P, VIRDEE, S, BEISHON, S. 1997. *Ethnic minorities in Britain: diversity and disadvantage.* London, Policy Studies Institute

MURRAY, U. and BROWN D. 1998 *They Look After Their Own, Don't They?* London, Social Services Inspectorate, Department of Health.

NAZROO, J. 1997 *The health of Britains Ethnic Minorities, Findings from a National Survey*. London: Policy Studies institute

NAZROO, J. 1998 'Genetic, cultural or socio-economic vulnerability? Explaining ethnic inequalities in health' *Sociology of Health and Illness* vol 20(5), 710-730.

NORMAN A (1985) *Triple Jeopardy: growing old in a second homeland.* Policy Studies in Ageing 3, Centre for Policy on Ageing

RAIT, G. and BURNS A. 1997 'Appreciating background and culture: the South Asian elderly and mental health' *International Journal of Geriatric Psychiatry*, vol 12, pp.973-77.

RITCH A, EHTISHAM M, GUTHRIE S, TALBOT J, LUCK M and TINSLEY R. 2001 'Ethnic influence on health and dependency of elderly inner city residents'.*Sociology of Health & Illness* Vol 23 Issue 6 pp 829

SHAH, A. 1998. 'The psychiatric needs of ethnic minority elders in the UK', *Age and Ageing* (3): 267(3).

SMAJE, C. 1995 *Health, Race and Ethnicity: Making Sense of the Evidence.*

London, King's Fund.

SMAJE, C. and LE GRAND J. 1997 'Ethnicity, equity and the use of health services

in the British NHS' *Social Science & Medicine* 45(3): 485(12).

WALKER R and AHMAD W, 1994 'Asian and Black elders and community care:

a survey of care providers' *New Community*, 20 (4) 635-646)

Table 1 Prevalence and standardised risk ratios of high GHQ12 score (4 or more), by ethnic group Health Survey for England 1999

	Black Caribbean	Indian	Pakistani	Bangladeshi	Chinese	Irish	General population
Men							
Observed %	16	16	18	26	3	18	15
Standardised risk ratio[1]	1.05	1.05	1.34	1.87	0.19	1.25	1
Women							
Observed %	23	23	22	23	8	20	19
Standardised risk ratio	1.22	1.26	1.27	1.57	0.39	1.12	1

[1] These show the 'risk' of a group experiencing a High GHQ12 score relative to the risk in the general population which is set at 1. As an example, the value of 1.87 for Bangladeshi men indicates this group is 87% more likely to have a High GHQ score (after allowing for age differences) than the population in general. Conversely the risk ratio of 0.19 for Chinese men indicate a prevalence 81% lower than in the general population.

Source: 1999 HSE (Table from Calderwood and Tait, 2001)

Table2: Differences in percentage of ethnic minority population with a 'High' GHQ12 score, using different recodings of ethnic group and age

Statistically significant differences (95%) are in **bold**
Italics signifies an estimate based on less than 30 cases

Population subgroup		Difference (percentage point) in estimated share with a 'High' GHQ12 score			
		Difference from General population in that age band		Difference from youngest age category (16-34) in that ethnic group	
		men	women	men	women
Indian	65+	+13.5	*-0.7*	*+17.2*	*-5.1*
	55+	+6.8	+6.5	+11.2	0
Pakistani	65+	*+19.8*	*+5.1*	*+18.5*	*+5.3*
	55+	+11.6	+1.4	+11.1	-0.5
Bangladeshi	65+	**+34.7**	***+41.6***	**+30.4**	***+46.2***
	55+	**+28.9**	**+32.1**	**+25.4**	**+34.6**
Pakistani/Bangladeshi	65+	**+30.3**	**+33.0**	**+27.8**	**+35. 3**
	55+	**+21.6**	**+19.6**	**+19.8**	+3.5
South Asian	65+	**+20.5**	+11.9	**+19.8**	+12.2
	55+	**+15.0**	**+12.8**	+11.2	**+11.1**

Source: HSE 1999: Individual file

Figure 1 Percent reporting 'High' GHQ12 Score, by sex, ethnic group and age
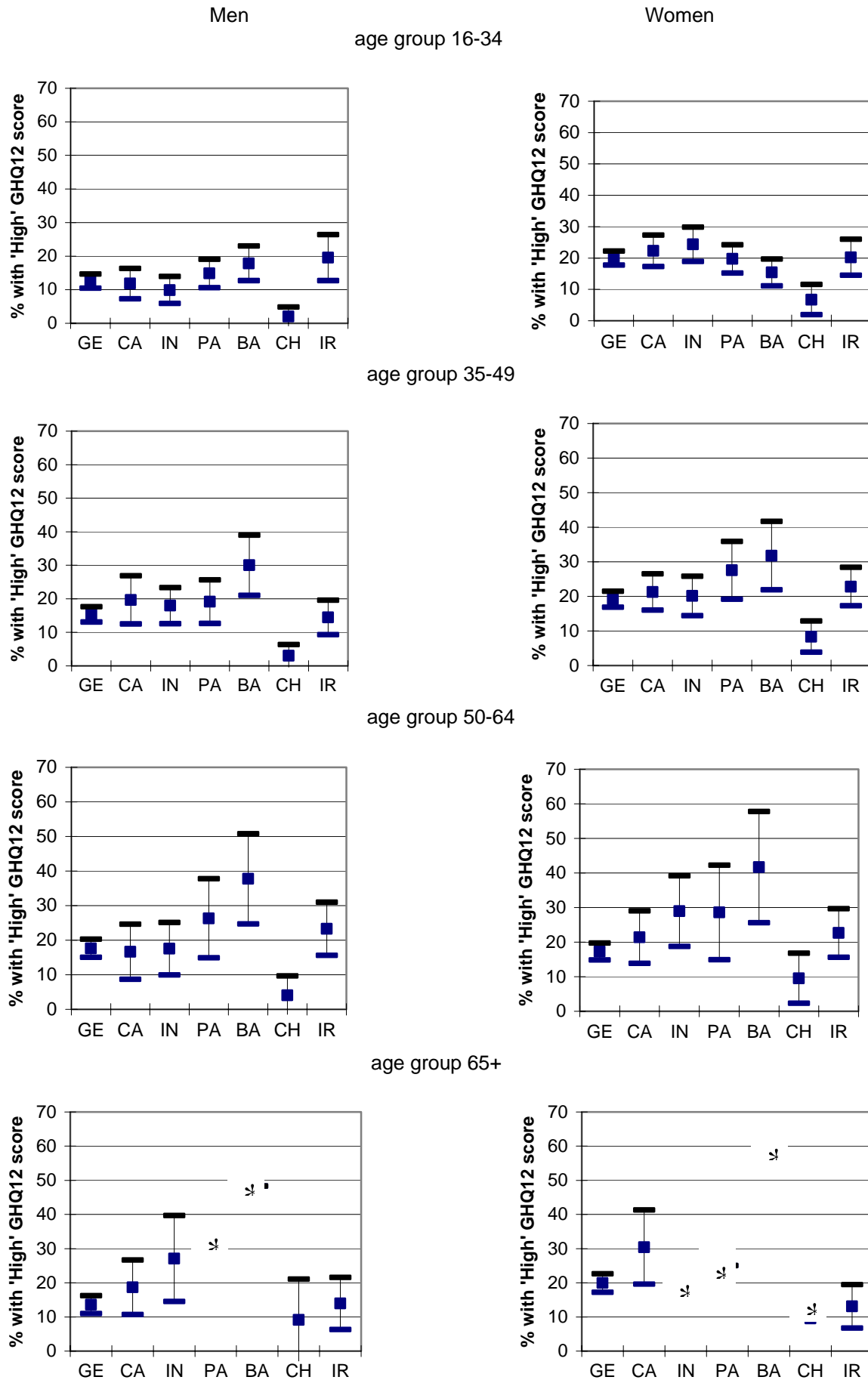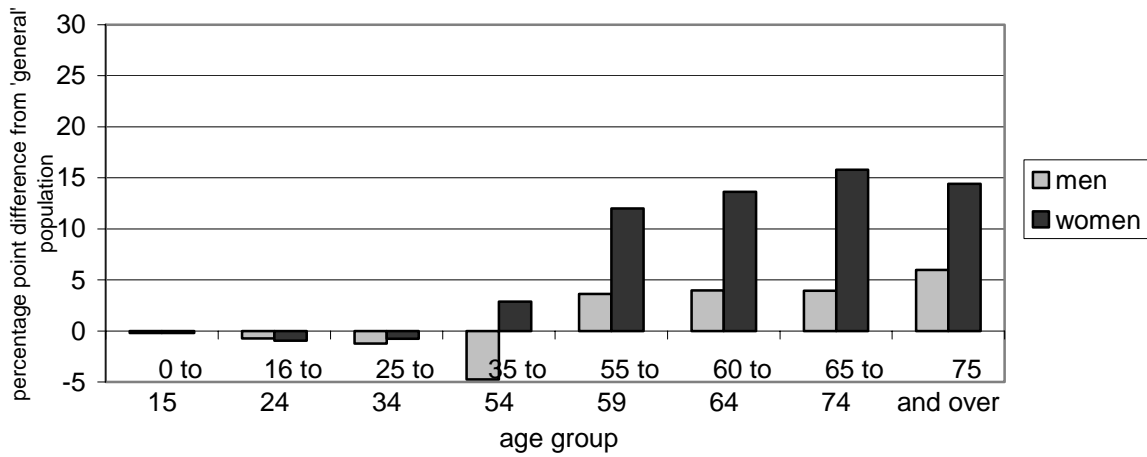
GE=General Population; CA=Black Caribbean; IN=Indian; PA=Pakistani; BA=Bangladeshi; CH=Chinese; IR=Irish
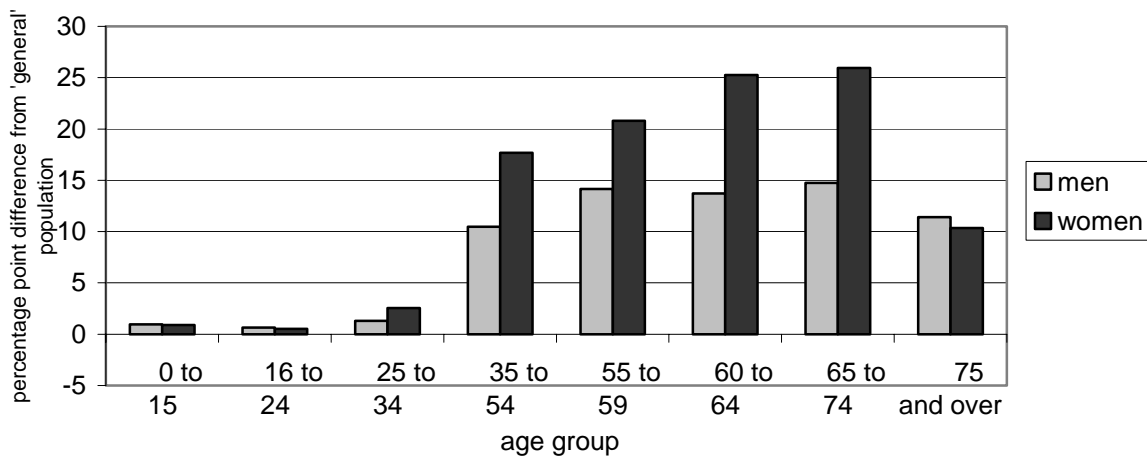
↯ = estimate based on less than 30 cases

Figure 2 Differences in percentage of ethnic group population with self-reported 'not good' general health compared to general population: by age and sex for selected ethnic groups, 2001 Census.

a) Indian

b) Pakistani

c) Bangladeshi