
Review article

Methods for epidemiological surveys of ethnic minority groups

Nishi Chaturvedi, Paul M McKeigue

Abstract

Objective – Research into the health of minority ethnic groups is often restricted by methodological difficulties. These include the lack of accurate population denominators, the choice of an appropriate sampling frame, correctly assigning ethnic group, and biases in techniques used for sampling and investigation. This article reviews the available sources of mortality and morbidity data, and assesses their uses and limitations for research involving ethnic minority groups. Suitable sampling frames and review methods used to assign ethnicity are discussed. Sources of bias are highlighted and methods used to overcome these biases are presented.

Criteria for inclusion of articles – Articles have been chosen which best illustrate the problems encountered and show how these problems can be addressed.

Conclusions – The increased documentation of ethnic origin on routine data sources is welcomed, but attention must be paid to ensuring that congruent definitions in data collection are used. The worrying consequences of the Commission of European Communities directive, which describes the need for explicit consent to be obtained from subjects before data is used for anything other than its original purpose, are discussed.

(*J Epidemiol Community Health* 1994;48:107-111)

Research on the health of migrants and ethnic minorities has helped to advance our understanding of disease aetiology^{1,2} and can also be helpful for the communities concerned. The disease rates of recent migrants usually reflect patterns in the country of origin, at least in the social stratum from which the migrants originated. Comparing the mortality of recent migrants with that in the general population can thus help to validate international differences in cause specific mortality. For instance, the low coronary mortality of Italian immigrants to England and Wales compared with the national average indicates that the difference in coronary mortality between Italy and the UK is unlikely to result from differences in certification practice.³ Where no data on cause specific mortality in the coun-

tries of origin are available, as in South Asia, the mortality of recent migrants to the UK from South Asia may yield clues to patterns of mortality in the countries of origin. Comparing the mortality of long settled migrants with rates in the host country and the country of origin can help to establish the relative contribution of genetic and environmental influences to international differences in mortality, and to determine the age at which lifetime risk is set.³ Studies of health care utilisation in ethnic minority groups have highlighted deficiencies in access to care and provided a basis for health care planning.⁴⁻⁶

Methodological difficulties arise in research on the health of ethnic minorities. These difficulties include obtaining population denominators, assembling sampling frames, assigning ethnic group, and minimising the biases related to sampling and data collection. This review will examine these difficulties and discuss techniques that can be used to overcome these problems.

Race or ethnicity?

This review is not intended to reopen the debate on the use of these terms; race is used to define a group which has in the past shared a common gene pool in which genetic differentiation from other populations has occurred, whereas ethnicity defines a population which shares cultural and linguistic characteristics.⁷ Ethnic groupings do not necessarily define homogeneous populations,⁸ but it is often useful to start with broad categories based on region of origin before attempting to define whether more specific categories are relevant to the health problems under study.

Ethnic minority populations in the UK

From the 1991 census it is estimated that 2.94 million (5.9%) of the total population of England and Wales belong to an ethnic minority group.⁹ Altogether 2.9% of the population are South Asian (Indian, Pakistani, or Bangladeshi) and 1.4% are Afro-Caribbean (black Caribbean) or black African. Afro-Caribbean and South Asian migration to the UK was stimulated by the post war demand for labour. For both groups, migration began in the early 1950s, peaking in the early 1960s for black Caribbeans, and in the late 1960s for South

Department of
Epidemiology and
Public Health,
University College
and Middlesex School
of Medicine, London
WC1E 6EA
N Chaturvedi

Department of
Epidemiology and
Population Sciences,
London School of
Hygiene and Tropical
Medicine, Keppel
Street, London
WC1E 7HT
P M McKeigue

Correspondence to:
Dr P M McKeigue.

Accepted for publication
24 July 1993

Asians.¹⁰ Africanisation policies in east Africa led to a further influx of South Asians after 1968.

Although it is possible that in some migrant groups, those who are unwell may prefer to return to the country of origin, there is no evidence that this is an important source of bias in studies of mortality and morbidity among ethnic minorities in the UK. Those who surmount the obstacles to international migration are generally more healthy than those who do not migrate from their native country. This "healthy migrant effect" tends to wear off with time.¹¹ People from ethnic minorities may be more likely than the general population to be working in low status occupations which do not match their educational qualifications. Occupational class as an indicator of socioeconomic status may thus be a misleading indicator of socioeconomic status when comparing ethnic groups.^{12,13} Housing tenure may be a more useful indicator, when adjusted for geographical location. Ethnic differences in tenure have been noted, particularly when South Asian and Europeans are compared, but these differences are attenuated when inner cities are examined, and when Afro-Caribbeans are compared with Europeans.¹³

Use of routine data

BIRTHS, DEATHS, AND CENSUS DATA

Country of birth is recorded on census records, death certificates, and birth registrations. Analyses of mortality by country of birth are thus possible, using either census estimates or birth registrations as the denominator. Census denominators are reliable only if analyses are based upon a time interval around the census. These analyses have been used to demonstrate ethnic differences in health and generate aetiological hypotheses.¹⁴⁻²² In people of black or Asian ethnicity who are now aged over 40 years (born before about 1953) almost all were born outside the UK, so that country of birth is still usable as an indicator of ethnicity in this age range. In those aged less than 40 years, the proportion of Afro-Caribbean or South Asian people who are born in the UK is now high. Other limitations of mortality analyses by country of birth are that published tables do not distinguish between west African (predominantly black African) migrants and east African (predominantly South Asian) migrants, and no data are recorded on age at migration.

For some population groups, mortality for migrants is usually calculated for an aggregate of several countries, for example the Indian subcontinent includes all those born in India, Pakistan, Bangladesh, and Sri Lanka, and Ireland includes both north and south Ireland. Calculation of separate rates for each of these countries may be misleading, as informant reporting of country of birth may be incomplete or inaccurate, or may ignore changes in geographical and political boundaries.

In the 1991 census a question on ethnic group was included for the first time. The

ethnic categories chosen were broad, so that "white" includes Irish and Cypriot migrants, who have different mortality and morbidity patterns than the general population.¹⁵ Self classification of ethnicity can be misleading, as respondents may classify themselves differently on repeat testing. It will not be possible to calculate mortality by ethnicity unless ethnicity is recorded on death certificates, which will require legislation.

ROUTINE MORBIDITY DATA

Cancer registrations include country of birth, and can thus be used with census denominators to calculate incidence rates by country of birth. Congenital malformation registrations include a unique identifier so that the case can be identified if necessary, but ethnicity is not included in the data held by the Office of Population Censuses and Surveys. Infectious disease notifications do not routinely include ethnicity, and to obtain the ethnic group of cases it is necessary to conduct special surveys, as in the current national survey of tuberculosis notifications.

Some other datasets can be used to examine morbidity rates without relying on census denominators. The Third National Morbidity Study, in which consultations during 1981 were linked to the census records of patients registered with 25 practices, included ethnicity defined by the country of birth of the head of household. The current National Morbidity Survey collects demographic data as part of the survey and includes ethnicity. The General Household Survey asks subjects to assign themselves to one of five categories: white, Indian, Pakistani, Afro-Caribbean, or other. The National Food Survey could be a useful source of data on dietary patterns but does not collect any data on ethnicity.

Sampling frames for surveys of ethnic groups

Surveys of the health of ethnic minorities depend on having a sampling frame for the groups under study. Census records cannot be used as a sampling frame for population based studies (for reasons of confidentiality), but census data can be used to identify geographical areas where sampling may provide a high proportion of subjects from groups under study.²³ Sampling fractions can be estimated for geographical strata, to take account of population dispersion. This selection of areas with concentrations of ethnic minority groups may lead to bias, because people who settle in areas where ethnic minority groups are concentrated are likely to be different from those who settle in areas where most residents are of native European origin.²⁴

FAMILY HEALTH SERVICES AUTHORITY LISTS AND OTHER HEALTH SERVICE REGISTERS

The lists of patients registered with general practitioners held by the family health services authority (FHSA) have been used to provide a

denominator population to examine ethnic differences in consultation,²⁵ and as a sampling frame for cross sectional epidemiological surveys,^{1,26,27} evaluation of health services,⁵ and assessing needs.⁴ It is estimated that over 98% of the population is registered with a GP.²⁸ Although the proportion of the population not registered with a GP is higher in inner cities (where a large proportion of people from minority ethnic groups live), even in inner London only 5% of the "white" and 2% of the "black" and "Asian" groups were not registered.²⁹ Although ethnic group is not recorded on FHSA lists, when preparing a list of patients for a survey it is usually possible for practice staff to assign ethnicity to patients on the list.

The most serious problem with sampling based on FHSA lists is wrong addresses. In inner city areas, where populations are highly mobile, the proportion of patients who are not living at the addresses given on the FHSA lists is often as high as 30%.^{30,31} Until recently there has been little effort to keep the information up to date. Surveys based on FHSA lists are thus likely to be biased towards those who change address less frequently, and this bias may not be the same in each ethnic group. Before starting a survey, it is often helpful to check the addresses on FHSA lists against manual or computerised records held by the practice, or against the electoral roll.³² Recorded delivery mailshots are a quick and cheap way of identifying most of the wrong addresses in the sampling frame. To obtain an accurate denominator for the response rate it is still necessary to visit the households where a recorded delivery letter has been accepted but no response received, to check that the subject still lives there.

Other possible sampling frames based on NHS records include child health registers and NHS numbers. Child health registers could be used to provide a sampling frame of mothers and young children but the content of these databases is not nationally standardised. NHS numbers are allocated either at birth or on immigration, and could potentially be used to identify migrants. But in practice they cannot be used for this purpose as new numbers are also given to others in adulthood, such as those who lose their original NHS number. One opportunity to use NHS numbers arose when Vietnamese refugees to the UK were allocated a block of consecutive NHS numbers. By flagging these records at the NHS Central Registry, it was possible to examine cancer morbidity and mortality in this group.³³

ELECTORAL REGISTER

The electoral register lists for each household the names of people aged over 18 years who are UK or Commonwealth citizens. Sampling from these registers has been used to identify South Asians for population surveys.^{11,24} This is inefficient if a relatively narrow age range is to be sampled, as the electoral register does not include age. In 1984 it was estimated that in

inner London around 14% of those eligible were not on the electoral register,³⁴ and it is speculated that the introduction of the community charge has further reduced the completeness of registration.³⁵ This bias can be dealt with by asking electors to enumerate their households which are missing from the electoral roll.²⁴

WORKFORCE SAMPLES

Some studies have used occupational groups to explore ethnic differences in disease.³⁶⁻³⁸ Sometimes this gives access to a group which is difficult to sample by other methods.³⁹ There may be some indication from employer's records of the relative proportions of subjects from each ethnic group. Another advantage is that it may be possible to match accurately for occupational status, which is usually difficult in community based sampling frames. A source of bias is that unemployment rates may be higher in ethnic minorities than in native populations.¹² Those in employment are selected for favourable health status, and this selection may be more stringent for ethnic minorities than for the native population. The most serious disadvantage of occupationally based samples is that it is now difficult to find large workforces on a single site in which ethnic minorities make up a high proportion. Current changes in employment, including economic factors, privatisation, and contracting out of services, mean that in future it will be even harder to find large stable workforces with a single employer.

TELEPHONE SAMPLING

Telephone sampling has been used extensively in the United States to draw population based samples of blacks and whites.^{40,41} Random digit dialling is preferred to sampling from telephone directories as it is not biased by unlisted numbers and out of date directories. Databases exist in the US which classify exchanges by ethnicity and income, so that sampling can be weighted towards areas with high proportions of minority ethnic groups. Unless the sample is weighted so that households are sampled in proportion to their size, telephone samples are biased towards single people and small households. In the US it is estimated that 93% of households possess a telephone, but the rates are much lower for blacks and other minority groups.⁴² In the UK only 80% of households possess a telephone and telephone sampling is not widely used.

SPECIAL CENSUS

In health surveys in developing countries it is usually necessary to undertake a special census to create a sampling frame for the population under study. The census may include the entire area under study, or a sample of blocks within the area. This method may be useful even in the UK for studies of groups such as

Bangladeshis who are concentrated in a few areas and for whom FHSA lists are inaccurate.

Assignment of ethnicity

Analysis of names has been used to allocate subjects to a particular ethnic group both for hospital based surveys^{43,44} and in general practice to provide community based sampling frames.¹⁴ This method is especially useful when studying South Asians, as Hindu and Sikh names are easily identifiable. It is possible to define groupings based on religion and region of origin, which are relevant to health related behaviour such as alcohol use and diet.¹⁶ Classification based on names is usually more valid than classification based on place of birth, as it identifies UK born South Asians, or people of European descent born abroad.⁴ It is not usually possible to distinguish the country of origin of people with Muslim names, and Indian Christians may be misclassified as European. Assignment of ethnicity based on names usually requires the skills of trained staff. A computer programme was developed to identify Chinese names,⁴⁵ but despite the use of a large referent database containing over 5000 Chinese names, 19% of new subjects could not be categorised.

For studies which rely on census denominators, it is preferable to ask subjects to identify their own ethnicity according to the categories used in the 1991 census question. In field surveys, it is unwise to rely on an interviewer's assessment of observation of appearance and skin colour. This may be a poorly valid marker of ethnicity, especially in subjects of mixed parentage. In the US, there have been attempts to use skin colour to define the degree of European admixture of those who are otherwise classified as black. Skin colour was found to be correlated with blood pressure, but confounding by socioeconomic status may have contributed to this association.⁴⁶ Combining an interviewer's assessment of ethnicity with parents' country of birth has been successfully used in several UK epidemiological studies of adult populations,^{1,27} but this method depends upon information about historical migration patterns which may not be available in other settings or with different age groups.

Non-response bias and information bias

Although response rates to invitations to participate in epidemiological research are generally as high for ethnic minorities as in the native population, among ethnic minority groups respondents are likely to be more literate and more fluent in English than non-respondents. In mailshots which include a covering letter from the subject's general practitioner, we have found that the relationship of patients with their general practitioner is the most important factor determining the response rate of any group. Literacy rates are not high in some migrant groups, especially in women and older subjects,⁴ and it is often necessary to follow mailshots with telephone calls or visits.

Future developments

Standardised recording of ethnicity on records such as death certificates, FHSA registers, and hospital admissions is necessary if these data sources are to provide meaningful information on the health of ethnic minorities. It is now planned to include ethnic origin on both the Hospital Episode Statistics (HES) and on FHSA databases, together with new NHS numbers as individual identifiers.⁴⁷ It is not clear how the assignment of ethnicity will be standardised against the census classification, which is a prerequisite for calculating rates free of numerator denominator bias. The Health Survey for England⁴⁸ will soon be expanded to cover 17 000 adults each year, so that about 1000 people from ethnic minorities will be sampled each year if the sample is representative. To monitor changes in the health of specific minority groups, it will be necessary to aggregate several years' data, or to conduct a separate but comparable survey for population groups of interest.

Many of the techniques discussed in this review of methods for studying the health of ethnic minorities will be illegal if the Commission of the European Union implements its draft directive on data protection in health related research.⁴⁹ This directive requires the consent of subjects before personal data can be used for any purpose other than that for which they were originally collected. It is of course difficult to contact subjects for consent when the sampling frame is being assembled, before the survey has begun. The future of epidemiological research in Europe depends upon persuading the Commission to withdraw this draft.

- 1 McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet* 1991;337:382-6.
- 2 Syme SL, Marmot MG, Kagan A, Kato H, Rhoads GG. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: introduction. *Am J Epidemiol* 1975;102:477-80.
- 3 Marmot MG, Adelstein AM, Bulusu L. *Immigrant mortality in England and Wales 1970-78*. OPCS Studies of Medical and Population Subjects no 47. London: HMSO, 1984.
- 4 Donaldson LJ. Health and social status of elderly Asians: a community survey. *BMJ* 1986;293:1079-82.
- 5 Bradley SM, Friedman EH. Cervical cytology screening: a comparison of uptake among 'Asian' and 'non-Asian' women in Oldham. *J Public Health Med* 1993;15:46-51.
- 6 Goldberg KG, Hartz AJ, Jacobsen SJ, Krakauer H, Rimm AA. Racial and community factors influencing coronary artery bypass graft surgery rates for all 1986 Medicare patients. *JAMA* 1992;267:1473-7.
- 7 Cooper R. A note on the biologic concept of race and its application in epidemiologic research. *Am Heart J* 1984;108:715-23.
- 8 Bhopal RS, Phillimore P, Kohli HS. Inappropriate use of the term 'Asian': an obstacle to ethnicity and health research. *J Public Health Med* 1991;13:244-6.
- 9 OPCS. *OPCS National Monitor: 1991 Census Great Britain*. London: HMSO, 1992.
- 10 Lomas GB. *Census 1971: The coloured population of Great Britain*. London: Runnymede Trust, 1973.
- 11 Williams R. Health and length of residence among South Asians in Glasgow: a study controlling for age. *J Public Health Med* 1993;15:52-60.
- 12 Ohri S, Faruqi S. Racism, employment and unemployment. In: Bhat A, Carr-Hill R, Ohri S, eds. *The Radical Statistics Race Group - Britain's black population*. Aldershot: Gower, 1988:61-100.
- 13 Brown C. *Black and white in Britain - the third PSI survey*. London: Heineman, 1984.
- 14 Adelstein AM, Marmot MG. Migrant studies in Britain. *Br Med Bull* 1984;40:315-19.
- 15 OPCS. *Mortality and geography: a review in the mid-1980s*. The Registrar-General's decennial supplement for England and Wales, series DS no 9. London: HMSO, 1990.
- 16 Balarajan R, Adelstein AM, Bulusu L, Shukla V. Patterns

- of mortality among migrants to England and Wales from the Indian subcontinent. *BMJ* 1984;**289**:1185-7.
- 17 Cochrane R. Mental illness in immigrants to England and Wales: an analysis of mental hospital admissions, 1971. *Soc Psychiatry* 1977;**12**:25-35.
 - 18 Balarajan R. Ethnic differences in mortality from ischaemic heart disease and cerebrovascular disease in England and Wales. *BMJ* 1991;**302**:560-4.
 - 19 Balarajan R, Raleigh VS, Botting B. Sudden infant death syndrome and postneonatal mortality in immigrants in England and Wales. *BMJ* 1989;**298**:716-20.
 - 20 Terry PB, Condic RG, Settartree RS. Analysis of ethnic differences in perinatal statistics. *BMJ* 1980;**281**:1307-8.
 - 21 Macfarlane A, Mugford M. *Birth counts: statistics of pregnancy and childbirth*. London: HMSO, 1984.
 - 22 Schoendorf KC, Hogue CJR, Klienman JC, Rowley D. Mortality among infants of black as compared with white college educated parents. *N Engl J Med* 1992;**326**(23):1522-6.
 - 23 Kalton G, Anderson DW. Sampling rare populations. *Journal of the Royal Statistical Society* 1986;**149**:65-82.
 - 24 Ecob R, Williams R. Sampling Asian minorities to assess health and welfare. *J Epidemiol Community Health* 1991;**45**:93-101.
 - 25 Gillam SJ, Jarman B, White P, Law R. Ethnic differences in consultation rates in urban general practice. *BMJ* 1989;**299**:953-7.
 - 26 Haines AP, Booroff A, Goldenberg E, Morgan P, Singh M, Wallace P. Blood pressure, smoking, obesity and alcohol consumption in black and white patients in general practice. *J Hum Hypertens* 1987;**1**:39-46.
 - 27 Chaturvedi N, McKeigue PM, Marmot MG. Resting and ambulatory blood pressure differences in Afro-Caribbeans and Europeans. *Hypertension* 1993;**22**:90-6.
 - 28 Ritchie J, Jacoby A, Bone M. *Access to primary health care: an enquiry carried out on behalf of United Kingdom health departments*. London: HMSO, 1981.
 - 29 Bone M. *Registration with general medical practitioners in inner London: a survey carried out on behalf of the Department of Health and Social Security*. London: HMSO, 1984.
 - 30 Fraser RC. Patient movements and the accuracy of the age-sex register. *J R Coll Gen Pract* 1982;**32**:615-22.
 - 31 Silman AJ. Age-sex registers as a screening tool for general practice: size of the wrong address problem. *BMJ* 1984;**289**:415-16.
 - 32 Bickler G, Sutton S. Inaccuracy of FHSA registers: help from electoral registers. *BMJ* 1993;**306**:1167.
 - 33 Swerdlow AJ. Mortality and cancer incidence in Vietnamese refugees in England and Wales: a follow-up study. *Int J Epidemiol* 1991;**20**:13-19.
 - 34 Hickman M. *Compiling the electoral register*. London: OPCS, 1991.
 - 35 Todd JE, Eldridge J. *Electoral registration in inner city areas, 1983-1984: a survey carried out by the Social Survey Division of OPCS on behalf of the Home Office*. London: HMSO, 1987.
 - 36 Cruickshank JK, Jackson SHD, Beevers DG, Bannan LT, Beevers M, Stewart VL. Similarity of blood pressure in Blacks, Whites and Asians in England: the Birmingham Factory Study. *J Hypertens* 1985;**3**:365-71.
 - 37 Meade TW, Brozovic M, Chakrabarti R, Haines AP, North WRS, Stirling Y. Ethnic group comparisons of variables associated with ischaemic heart disease. *Br Heart J* 1978;**40**:789-95.
 - 38 Fontbonne A, Papoz L, Eschwege E, Roger M, Saint-Paul M, Simon D. Features of insulin-resistance syndrome in men from French Caribbean islands. *Diabetes* 1992;**41**:1385-9.
 - 39 Watt IS, Howel D, Lo L. The health care experience and health behaviour of the Chinese: a survey based in Hull. *J Public Health Med* 1993;**15**:129-36.
 - 40 Waksberg J. Sampling methods for random digit dialing. *Journal of the American Statistical Society* 1978;**73**:40-6.
 - 41 McGraw SA, McKinlay JB, Crawford SA, Costa LA, Cohen DL. Health survey methods with minority populations: some lessons from recent experience. *Ethnicity of Disease* 1992;**2**:273-87.
 - 42 Thornberry TO, Massey JT. Trends in United States telephone coverage across time and subgroups. In: Groves RM, Biemer PB, Lyberg LE, et al, eds. *Telephone survey methodology*. New York: Wiley & Sons, 1988:25-49.
 - 43 Nicoll A, Basett K, Ulijaszek SJ. What's in a name? Accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations. *J Epidemiol Community Health* 1986;**40**:364-8.
 - 44 Jayanthi V, Probert CSJ, Pinder D, Wicks ACB, Mayberry JF. Epidemiology of Crohn's disease in Indian migrants and the indigenous population in Leicestershire. *Q J Med* 1992;**82**:125-38.
 - 45 Coldman AJ, Braun T, Gallagher RP. The classification of ethnic status using name information. *J Epidemiol Community Health* 1988;**42**:390-5.
 - 46 Boyle Jr E. Biological patterns in hypertension by race, sex, body height, and skin color. *JAMA* 1970;**213**:1637-43.
 - 47 NHS Management Executive. *IM&T Strategy Overview*. London: Department of Health, 1992:1-19.
 - 48 Markowe H. The work of the Central Health Monitoring Unit in the Department of Health (England). *J Epidemiol Community Health* 1993;**47**:6-9.
 - 49 Commission of the European Communities. *Biomedical and health research in the European Community*. Luxembourg: Commission of the European Communities, Directorate XII-F-6, Medical Research Division, 1991.